Iranian Journal of Field Crops Research

Optimization and Validation of two Machine Learning Algorithms for Accurate Prediction of Irrigated Wheat Yield in Khorasan Razavi Province

Mohsen Jahan

Department of Agrotechnology, Faculty of Agriculture, Ferdowsi University of Mashhad (FUM)
Jahan@ferdowsi.um.ac.ir

DOI: 10.22067/jcesc.2025.93323.1395

Introduction

This study undertook a detailed comparison of two supervised machine-learning algorithms—Random Forest (RF) and eXtreme Gradient Boosting (XGBoost)—to predict irrigated wheat yield across 20 counties in Razavi Khorasan Province. Both models were trained on 70 % of the dataset (years 1383–1402) and tested on the remaining 30 %. Hyperparameter tuning was performed via a grid search coupled with five-fold cross-validation.

Materials and Methods

The two algorithms were first trained and optimized using 70 % of the data (≈14 years of data per county) and then tested on the remaining 30 % (≈6 years). Their hyperparameters were tuned via grid search with five-fold cross-validation. Hyperparameter Tuning was performed using Grid search over key parameters (e.g., number of trees n_estimators, maximum depth max_depth, learning rate for XGBoost) with 5-fold cross-validation. Afterwards, both models were evaluated and validated using RMSE (Root Mean Squared Error), R^2 (Coefficient of Determination), MAE (Mean Absolute Error), Willmott's d (Index of Agreement). Two machine learning algorithms, Random Forest and XGBoost, were developed to predict wheat performance at the district level. The performance values were initially predicted numerically using regression and then divided into three distinct classes using statistical percentiles: Low: 0th to 33rd percentile; Medium: 33rd to 66th percentile; High: 66th to 100th percentile. This classification was based on both the actual and predicted values for each district. The results of this classification are presented in the form of a confusion matrix for each model. Using three indices—Aridity Index, Seasonal Intensity of Temperature, and Growing Degree Days—districts in the province were clustered into three climatic zones. Then, wheat performance in these zones was analyzed based on two models. In this study, the machine learning algorithms Random Forest and XGBoost were implemented in Python 3.3 using the Scikit-learn library. DataFrame preparation and the clustering of the province's counties into climatic zones were carried out using R 4.3.2 and ArcGIS 10.8.2.

Results and Discussion

Alghoritms Performance: On average, RF reduced prediction error by ~19 % (395 vs. 492 kg/ha RMSE) and achieved a slightly higher agreement with observed yields (d=0.467 vs. 0.419). Random forest showed Best RF Performance for Khalilabad (RMSE = 141.19 kg/ha, MAE = 125.68 kg/ha, d = 0.37) and Bardeskan (RMSE = 187.69 kg/ha, d = 0.80); Worst RF Performance was obtained for Quchan (RMSE = 667.65 kg/ha, d = 0.36) and Torbat-e Heydarieh (RMSE = 581.33 kg/ha, d = 0.47). Best XGBoost Performance was obtained for Torbat-e Jam (RMSE = 269.36 kg/ha, d = 0.77), Neyshabur (RMSE = 377.91, d = 0.78). Worst XGBoost Performance resulted for Quchan (RMSE = 943.99 kg/ha, d = 0.25) and Torbat-e

Heydarieh (RMSE = 786.20 kg/ha, d = 0.39). In 6 out of 20 counties (Khaf, Mahvelat, Kalat-e Nader, Kashmar, Chenaran, Fariman) both RF and XGBoost performed nearly identical errors (ΔRMSE < 15 kg/ha), indicating similar predictive power under those local conditions.
Feature Importance: Daily Minimum Temperature (Tmin): Ranked #1 in RF's importance list; ranked #3 in XGBoost. Seasonal Tmin (TminGS): Consistently #3 in both models. Other Key Predictors: Precipitation over the growing season (Prec, PGS), Growing Degree Days (GDDGS), and Evapotranspiration (ETGS) all contributed substantially, though with 4th–8th ranks markedly lower in RF than in XGB.
Classifiction of Yield into Three Performance Classes: Using percentile thresholds—Low (0–33rd), Medium (34–66th), High (67–100th)—the models were also evaluated as classifiers. Low-Performing Counties Needing Intervention Seven counties (35 % of the sample)- Quchan, Torbat-e Heydarieh, Sarakhs, Kalat-e Nader, Gonabad, Neyshabur, Taybad- fell in the Low performance class across both models. These areas should be prioritized for targeted agronomic management and resource allocation.
Cluster-Based Insights: Counties were grouped into three agro-climatic clusters: Very Dry & Hot (4 counties): RF outperformed XGB in all (e.g., Bardeskan, Khalilabad, Mahvelat, Sarakhs). Semi-humid Cooler (6 counties): Mixed results-RF won in 4 (Chenaran, Kalat-e Nader, Mashhad, Nishapur); XGB was slightly better in 2 (Fariman, Quchan). Warm Semi-arid (10 counties): RF superior in 7; equivalence in Taybad & Torbat-e Heydarieh; XGB edged ahead only in no counties here.
Conclusion: Overall, the Random Forest model showed better results for predicting wheat yield in Razavi Khorasan province, especially in most counties. Although the XGBoost model has higher potential for modeling complex patterns, Random Forest performed more accurately in conditions of greater data dispersion. In conclusion, the results of this study emphasize that by utilizing climatic and agricultural data, machine learning algorithms can be optimized to not only achieve high accuracy but also provide a clear analysis of the role of each variable in wheat yield.

KeyWords: Random Forest, XGBoost, Grid Search, Cross-Validation, Yield Class.

Ferdowsi University of Mashhad

Iranian Journal of Field Crops Research

شماره ۷۱۲۵
تاریخ ۱۴۰۴/۰۴/۳۱
پیوست ندارد

نشریه علمی
پژوهشهای زراعی ایران

باسلام بدینوسیله به اطلاع می‌رساند مقاله با عنوان:

بهینه‌سازی و اعتبارسنجی دو الگوریتم یادگیری ماشین برای پیش‌بینی دقیق عملکرد گندم آبی در استان خراسان رضوی

نویسنده: محسن جهان

در نشریه پژوهشهای زراعی ایران پس از بررسی، با توجه به نظر ارزیابان و تأیید ارزیاب نهایی، در جلسه مورخ **۱٤۰٤/۰٤/۳۱** به‌عنوان مقاله **پژوهشی** تصویب و در یکی از شماره‌های آتی براساس اولویت و با شناسه **DOI: 10.22067/jcesc.2025.93323.1395** چاپ خواهد شد.

سردبیر نشریه پژوهشهای زراعی ایران

نشریه پژوهشهای زراعی ایران
(قطب علمی گیاهان زراعی ویژه)

مشهد، میدان آزادی، پردیس دانشگاه فردوسی مشهد، دانشکده کشاورزی، دبیرخانه نشریات علمی
صندوق پستی : ۱۱٦۳
کد پستی: ۹۱۷۷۵
تلفن: ۳۸۸۰٤٦۱۹-۰۵۱
نمابر :۳۸۷۸۷٤۳۰-۰۵۱
cesc@um.ac.ir

Iranian Journal of Field Crops Research

بهینه‌سازی و اعتبارسنجی دو الگوریتم یادگیری ماشین برای پیش‌بینی دقیق عملکرد گندم آبی در استان خراسان رضوی

چکیده

این تحقیق به مقایسه دو الگوریتم یادگیری ماشین، رندوم فارست و ایکس‌جی‌بوست، در پیش‌بینی عملکرد گندم آبی در ۲۰ شهرستان استان خراسان رضوی پرداخته است. دو الگوریتم ابتدا با ۷۰ درصد داده ها آموزش دیده و بهینه سازی شدند و سپس با ۳۰ درصد داده‌ها تست شدند. بهینه‌سازی دو الگوریتم به کمک جستجوی شبکه ای با اعتبارسنجی متقابل پنج مرحله‌ای انجام گرفت. در ادامه، هر دو الگوریتم ارزیابی و تعیین اعتبار شدند. از نظر معیارهای ارزیابی، مانند جذر میانگین مربعات خطا (RMSE)، ضریب تبیین (R^2)، میانگین قدر مطلق خطا (MAE) و شاخص توافق ویلموت (d)، مدل رندوم‌فارست در بسیاری از شهرستان‌ها عملکرد بهتری داشت. در حالی که ایکس‌جی‌بوست در مناطقی با توزیع پیچیده‌تر داده‌ها (مانند مشهد) توانست به‌طور مؤثرتری عمل کند. مقایسه نتایج مدل‌ها در شهرستان‌های مختلف نشان می‌دهد که مدل رندوم فارست در بسیاری از شهرستان‌ها دقت بالاتری دارد. برای مثال، در شهرستان‌هایی مانند خلیل آباد، سبزوار، تربت‌جام، رشتخوار، نیشابور و گناباد، مدل رندوم‌فارست خطای کمتری نسبت به ایکس‌جی‌بوست دارد. از سوی دیگر، مدل ایکس‌جی‌بوست در بعضی مناطق خاص مانند مشهد، بردسکن، درگز، سرخس و تایباد عملکرد بهتری از خود نشان داد، که احتمالاً به ویژگی‌های خاص این شهرستان‌ها مربوط می‌شود. برای شهرستان‌های خاف، مه‌ولات و کلات نادر، هر دو مدل یکسان عمل کردند. در هر دو مدل، دمای کمینه روزانه نقش برجسته‌ای در پیش‌بینی عملکرد گندم ایفا کرده است. در مدل رندوم فارست، دمای کمینه روزانه رتبه اول را دارد، در حالی که در مدل ایکس‌جی‌بوست این ویژگی به‌طور معمول از اهمیت کمتری برخوردار است. سایر ویژگی‌ها مانند بارندگی و درجه روز رشد نیز به‌شکل قابل توجهی در دقت مدل‌ها تأثیرگذار بودند. مدل‌ها همچنین در طبقه‌بندی عملکرد گندم با استفاده از ماتریس درهم‌ریختگی ارزیابی شدند. مدل ایکس‌جی‌بوست به‌طور کلی دقت بالاتری در طبقه‌بندی کلاس‌های عملکرد داشت، در حالی که رندوم فارست در شبیه‌سازی دقیق‌تر داده‌های واقعی و کاهش خطاهای پیش‌بینی موفق‌تر عمل کرد. به‌طور کلی، مدل رندوم‌فارست برای پیش‌بینی عملکرد گندم در استان خراسان رضوی نتایج بهتری در بیشتر شهرستان‌ها نشان داد. اگرچه مدل ایکس‌جی‌بوست پتانسیل بالاتری برای مدل‌سازی الگوهای پیچیده دارد، رندوم فارست در شرایط پراکندگی بیشتر داده‌ها دقیق‌تر عمل کرد. نتایج این مطالعه تأکید کرد که با بهره‌گیری از داده‌های اقلیمی و زراعی، می‌توان الگوریتم های یادگیری ماشین را بهینه کرده و علاوه بر دقت بالا، تحلیلی شفاف و کاربردی از نقش هر متغیر در عملکرد گندم ارائه نمود.

واژه‌های کلیدی: رندوم‌فارست، ایکس‌جی‌بوست، جستجوی شبکه‌ای، تعیین اعتبار متقابل، کلاس عملکرد.

مقدمه

افزایش بی‌سابقه جمعیت جهان (پیش‌بینی شده ۹/۷ میلیارد نفر تا سال ۲۰۵۰) همراه با کاهش منابع آب و خاک و تشدید تغییرات اقلیمی، فشار مضاعفی بر نظام تولید مواد غذایی وارد می‌آورند و امنیت غذایی را به یکی از مهم‌ترین چالش‌های قرن بیست‌ویکم تبدیل کرده است (Aslan et al., 2024; Belete, 2025). در این شرایط، صرفاً افزایش تولید کافی نیست؛ بلکه بهینه‌سازی مصرف نهاده‌ها و بهبود بهره‌وری، کاهش ضایعات و پیش‌بینی دقیق عملکرد محصولات کشاورزی نیز از اولویت‌های اساسی سیاست‌گذاران و پژوهشگران است (Arun and Ghimire, 2019; Jhajharia et al., 2023). گندم (*Triticum*

(*aestivum* L.)، مهم‌ترین غله جهان، با سهم حدود۲۱درصد از تولید کالری جهانی، یکی از محورهای اصلی تأمین غذا محسوب می‌شود. برآورد می‌شود که تا سال ۲۰۵۰، تقاضای جهانی برای گندم تا ۶۰ درصد افزایش یابد، این در حالی است که نوسانات اقلیمی روند کاهشی عملکرد سالانه این محصول را تشدید می‌کنند(Nelson et al., 2009; Curtis and Halford, 2014). تغییرات اقلیمی تأثیر عمیقی بر عملکرد محصولات زراعی دارند. دما و بارش به‌تنهایی حدود ۳۰ تا ۴۰درصد از واریانس تغییرات عملکرد سالیانه را توضیح می‌دهند(Ray et al., 2015). اثر منفی افزایش دما بر عملکرد گندم، جو و ذرت در مقیاس جهانی به اثبات رسیده است. گزارش‌های هیئت بین‌الدول تغییر اقلیم[1] در چرخه ارزیابی هفتم (AR7) و به‌روزرسانی‌های منتشرشده پس از گزارش ششم ارزیابی (AR6) در سال ۲۰۲۳ منتشر شده است، نشان می‌دهد که گرمایش جهانی تا سال ۲۰۲۰ به ۱.۱ درجه سانتی‌گراد بالاتر از سطح پیش‌صنعتی رسیده است و پیش‌بینی می‌شود تا سال ۲۰۳۰، حتی در سناریوهای خوش‌بینانه، این افزایش ادامه یابد و به بیش از ۱.۵ درجه سانتی‌گراد نزدیک شود یا از آن فراتر رود (Assenget al., 2011; IPCC, 2022). وقوع و تکرار پدیده‌های شدید مانند خشکی‌های طولانی‌مدت، سیل و سرمازدگی، به‌خصوص در مناطق خشک و نیمه‌خشک، چالش‌های عدیده‌ای را برای تولید گندم به‌وجود آورده است که می‌تواند به‌سرعت زیربنای کشاورزی را تضعیف کند(Asseng et al., 2011; Raymundo et al., 2018). مدل‌سازی‌های اقلیمی آینده نیز کاهش عملکرد محصولات زراعی را پیش‌بینی و تأیید می‌کنند.

در ایران، مطالعات متعددی کاهش میزان بارش‌های فصلی، به ویژه در فصول بهار و پاییز، را گزارش کرده‌اند که به نوبه خود بحران کم‌آبی و تنش گرمایی را تشدید نموده است؛ همچنین، مدل‌سازی‌های اثر تغییر اقلیم بر سیستم های زراعی کشور در سال های گذشته نشان‌دهنده روند کاهشی در عملکرد گندم بوده و پیش‌بینی های آنها مبنی بر این که تا افق ۱۴۰۰خورشیدی، کاهش عملکرد در برخی مناطق تا ۴۸ درصد نیز می‌رسد، تحقق یافته است(Valizadeh et al., 2014; Koocheki and Nassiri-Mahallati, 2016; Boskabadi et al., 2022). نوسانات مکانی و زمانی عملکرد گندم در سطح جهان و ایران نه تنها ناشی از تغییرات اقلیمی، بلکه نتیجه تفاوت در دسترسی به منابع آبی، مدیریت زراعی و ادوات کشاورزی است (Bannayanet al., 2018; Roell et al., 2020). به عنوان مثال، در یک مطالعه پنلی در سه شهرستان استان خراسان رضوی، بیش از ۴۰ درصد تغییرات عملکرد گندم با متغیرهای اقلیمی توضیح داده شد (Bannayan et al., 2018).

پیش‌بینی دقیق عملکرد به کشاورزان کمک می‌کند تا منابعی مانند آب، کود و نیروی کار را به‌طور مؤثرتری به کار گیرند. پیش‌بینی عملکرد محصول از قبل، به برنامه‌ریزی بازاریابی و توزیع نیز کمک می‌کند و به ذینفعان – از کشاورزان گرفته تا تولیدکنندگان مواد غذایی و خرده‌فروشان – اجازه می‌دهد تا برای نوسانات عرضه و تقاضا، کاهش ضایعات و اطمینان از در دسترس بودن مواد غذایی آمادگی پیدا کنند (Aslan et al., 2024). با پیچیده‌تر شدن روابط میان متغیرهای اقلیمی، خاکی و مدیریتی، نیاز به روش‌های پیشرفته‌ای برای مدل‌سازی و پیش‌بینی عملکرد گندم احساس می‌شود. اگر چه امروزه مدل های فرآیند–محور با امکانات گسترده ای در دسترس هستند، اما نیاز به تنظیم پارامترها، واسنجی و اعتبارسنجی این مدل ها تحت شرایط خاص، و در نهایت وجود خطا در خروجی ها، همچنان موضوعی چالش برانگیز می باشد (Dhillon et al., 2023; Krishnadoss and Ramasamy, 2024). یادگیری ماشین به‌عنوان ابزاری نوظهور، به‌ویژه الگوریتم جنگل تصادفی[2] که با ترکیب نتایج چندین درخت تصمیم، توانایی بالایی در شناسایی الگوهای غیرخطی و چندمتغیری دارد، در سال‌های اخیر محبوبیت یافته است (Nayak et al., 2022; Jhajharia et al., 2023; Khodjaev et al., 2024). مطالعات متعددی گزارش کرده‌اند که

[1] **Intergovernmental Panel on Climate Change,**
[2] Random Forest

مدل‌های مبتنی بر جنگل تصادفی، نسبت به مدل‌های آماری کلاسیک، مانند رگرسیون خطی چندگانه، به‌طور محسوسی عملکرد بهتری در پیش‌بینی عملکرد گندم ارائه می‌دهند. برای مثال، جذر میانگین مربعات خطا[1] مدل رندوم فارست در یک مطالعه منطقه‌ای ۱۶۳/۹۰کیلوگرم بر هکتار (عملکرد گندم زمستانه) گزارش شد که در مقایسه با مدل رگرسیون خطی که ۶۵۳/۳۹ کیلوگرم بر هکتار بود، حاکی از دقت بیشتر آن می باشد (Belete, 2025). افزون بر این، روش‌های نوینی مانند رندوم فارست خودآموز[2] با امکان گسترش مجموعه‌های داده برچسب‌دار و بهبود دقت پیش‌بینی، نشان داده‌اند که می‌توان بدون نیاز به داده‌های گسترده اولیه، دقت مدل در پیش بینی عملکرد گندم زمستانه را تا حدود ۱۰ تا ۱۵درصد ارتقاء داد (Shen et al., 2024). در پژوهشی دیگر، به کارگیری پهپاد در برداشت داده ها از مزرعه گندم به منظور تهیه دیتاست برای پیش بینی عملکرد گندم، نشان داد که مدل ماشین بهینه سازی شده در مقایسه با مدل های خطی سنتی، دقت پیش بینی را به میزان ۷ تا ۱۲ درصد افزایش داد (Khodjaev et al., 2024). همچنین، پژوهش‌های اخیر بر اهمیت تنظیم اَبَرپارامترها با استفاده از تکنیک‌هایی مانند بهینه‌سازی شبکه‌ای[3] و آزمون های پَسا آزمون [4] مانند کراس ولیدیشن[5] تأکید دارند که منجر به بهبود ۷ تا ۱۲ درصدی دقت مدل‌های جنگل تصادفی می‌شود. جنگل تصادفی به دلیل توانایی بالای خود در مدل‌سازی روابط غیرخطی، مقاومت در برابر بیش‌برازش و ارائه شاخص‌های اهمیت متغیرها، به‌طور گسترده‌ای در مطالعات کشاورزی مورد استفاده قرار گرفته است (Roell et al., 2020; Pang et al., 2022).

الگوریتم دیگری که در یادگیری ماشین و مسائل رگرسیون و کلاس بندی، توانایی بالایی دارد و به عنوان یک مدل قدرتمند در پیش بینی هایی همچون موضوع مقاله حاضر به شمار می رود، الگوریتم ایکس‌جی‌بوست[6] است. ایده اصلی در این الگوریتم این است که به جای ساخت یک مدل بزرگ، چند مدل ساده (درخت تصمیم) را پشت‌سرهم می‌گذارد، به طوری که هر مدل بعدی، خطای مدل قبلی رو جبران می‌کند. هر دو الگوریتم از خانوادهٔ درخت تصمیم[7] بوده و در عمل برای اهداف مشابه به کار گرفته می شوند. خلاصه ای از توانایی ها و مقایسه این دو الگوریتم در جدول ۱ ارائه شده است.

جدول ۱- ویژگی های دو الگوریتم یادگیری ماشین رندوم فارست و ایکس‌جی‌بوست

Table 1 – Comparison of the Characteristics of Random Forest and XGBoost Algorithms

ویژگی/الگوریتم Feature / Algorithm	رندوم فارست Random Forest	ایکس‌جی‌بوست XGBoost
نحوه یادگیری Learning Approach	مدل ها به صورت موازی Bagging ساخته می شوند (نوع و ترکیب مدل ها از یکدیگر مستقل هستند) Models are built in parallel using Bagging (each model is independent)	مدل‌ها به‌صورت زنجیره‌ای Boosting ساخته می‌شوند (هر مدل، خطای مدل قبلی رو اصلاح می کند) Models are built sequentially using Boosting (each model corrects the errors of the previous one)
سرعت یادگیری Training Speed	روی CPU سریع Fast on CPU	روی GPU خیلی سریع Very fast on GPU
دقت نهایی Final Accuracy	خوب Good	معمولاً بهتر، مخصوصاً روی داده های پیچیده

[1] RMSE
[2] Self-Training Random Forest
[3] Grid Search
[4] Post Hoc Tests
[5] Cross Validation
[6] XGBoost (Extreme Gradient Boosting)
[7] Decision Tree

		Usually better, especially on complex data
کنترل روی بیش برازش **Overfitting Control**	خوب Good	عالی (به دلیل داشتن پارامترهایی مثل regularization) Excellent (thanks to parameters like regularization)
قابلیت کار با داده های ناقص Handling Missing Data	ندارد Not supported	دارد Supported

استان خراسان رضوی، با اختصاص بیش از ۴۸ درصد از اراضی کشاورزی به گندم آبی، یکی از مهم‌ترین قطب‌های تولید این محصول در ایران است. داده‌های آماری مرکز جهاد کشاورزی نشان می‌دهد که در سال زراعی ۱۴۰۰-۱۳۹۹، کاهش بارندگی تا ۳۸ درصد نسبت به متوسط بلندمدت، به همراه سرمازدگی، منجر به خسارت‌های گسترده در بیش از ۴۷ هزار هکتار از اراضی آبی و بیش از ۹۶ هزار هکتار از اراضی دیم شد (سازمان تحقیقات کشاورزی،۱۴۰۱). این شرایط، از یک سو، نشان از آسیب‌پذیری بالای این زیست‌بوم کشاورزی در برابر تغییرات اقلیمی دارد و از سوی دیگر، تنوع میکروکلیما و تفاوت مدیریتی گسترده در سطح شهرستان های این استان، ضرورت بررسی عمیق عوامل مؤثر بر عملکرد گندم آبی و مطالعه عمیق چگونگی تأثیر متغیرهای اقلیمی بر آنها را دو چندان می کند.

با توجه به این که، نوسانات چشمگیر بارش، تلفات گسترده در سال‌های پرتنش و کاهش شدید عملکرد طی دوره زمانی ۱۴۰۲-۱۳۸۳در برخی شهرستان‌ها نیازمند تحلیل‌عمیق‌تری است، این پژوهش، با هدف اصلی شناسایی و تحلیل روند بلندمدت عملکرد گندم آبی در ۲۰ شهرستان خراسان رضوی و توسعه مدل‌های پیش‌بینی دقیق مبتنی بر یادگیری ماشین» طراحی و اجرا شد. برای این منظور، مجموعه‌ای از شاخص‌های اقلیمی و زراعی مؤثر بر عملکرد گندم از پایگاه های داده و منابع علمی منتشر شده استخراج و محاسبه گردید. سپس، مدل‌های پیش‌بینی عملکرد گندم مبتنی بر دو الگوریتم رندوم فارست و ایکس‌جی‌بوست، برای هر شهرستان آموزش و توسعه داده شد و سپس دقت و اعتبار آنها مورد ارزیابی قرار گرفت. در آخر، اهمیت نسبی هر متغیر تعیین و راهکارهای مدیریتی سازگار با تغییرات آب‌وهوایی ارائه گردید. نتایج این مطالعه می‌تواند مبنایی برای طراحی سیاست‌های مدیریتی سازگار با تغییرات اقلیمی و بهینه‌سازی تولید گندم در خراسان رضوی فراهم آورد.

مواد و روش‌ها

منطقه مورد مطالعه

منطقه مورد مطالعه در این تحقیق، استان خراسان رضوی با مساحتی حدود ۱۱۸۸۵۴ کیلومتر مربع است که در مختصات جغرافیایی۳۳ درجه و۵۲ دقیقه تا ۳۷ درجه و۴۲ دقیقه عرض شمالی و۵۶ درجه و۱۹دقیقه تا ۶۱ درجه و۱۶ دقیقه طول شرقی قرار دارد. این استان دارای اقلیم متنوعی از نیمه‌خشک تا خشک کوهستانی بوده و در ارتفاع متوسطی حدود۹۰۰ تا۱۲۰۰متر از سطح دریا واقع شده است. این پژوهش در سطح ۲۰ شهرستان استان خراسان رضوی شامل مشهد، نیشابور، سبزوار، تربت حیدریه، کاشمر، قوچان، تربت جام، تایباد، گناباد، درگز، سرخس، فریمان، خواف، رشتخوار، بردسکن، چناران و بجستان، خلیل آباد، کلات نادر و مه‌ولات صورت گرفته است.

داده‌ها و اطلاعات مورد نیاز

اطلاعات اقلیمی: داده‌های روزانه دما (حداکثر، حداقل و میانگین)، بارندگی، تبخیر-تعرق، شاخص فصلی بودن درجه حرارت، درجه روز رشد و شاخص خشکی برای هر شهرستان در بازه زمانی۱۳۸۳ تا ۱۴۰۲ از ایستگاه های هواشناسی هر شهرستان تهیه شد و

در موارد اندکی که خلاء داده وجود داشت، و یا داده ها ناقص بودند، به منظور تکمیل و بازسازی داده ها، از پایگاه‌های معتبر و اعتبارسنجی شده جهت شبیه سازی داده های هواشناسی در سطح استان، نظیر APHRODITE, ERA5, IMERG, AgMERRA استفاده و در قالب فایل‌های سالانه پردازش شدند (Yaghoubi et al; 2022; Farhadi et al., 2023; Soltani et al., 2025). هر شهرستان بر اساس داده‌های آب و هوایی خود و موقعیت جغرافیایی در یک فایل جداگانه ذخیره شده و سپس با داده‌های عملکرد گندم ادغام شد.

اطلاعات عملکرد: داده‌های عملکرد گندم در واحد کیلوگرم بر هکتار برای هر شهرستان طی بازه زمانی ذکر شده از پایگاه آمار و اطلاعات سازمان جهاد کشاورزی استان خراسان رضوی، مراجعه حضوری به سازمان جهاد کشاورزی و سایر منابع معتبر گردآوری شد. بر اساس این داده‌ها و نیز برخی منابع علمی منتشر شده، مجموعه‌ای از متغیرهای مهم اقلیمی و زراعی استخراج و محاسبه شدند که به‌عنوان متغیرهای مستقل در مدل مورد استفاده قرار گرفتند (Farhadi et al., 2023). این متغیرها در جدول ۲ آورده شده اند.

جدول ۲- متغیرهای اقلیمی و زراعی مورد استفاده در الگوریتم های پیش بینی عملکرد گندم در استان خراسان رضوی

Table 2 - Climatic and Agronomic Variables Used in Wheat Yield Prediction Algorithms in Khorasan Razavi Province

متغیر اقلیمی/ زراعی	Climatic Variable
میانگین دمای روزانه در طول فصل رشد (درجه سانتی گراد) *Average daily temperature during the growing season (°C)*	Tmean (°C)
دمای بیشینه روزانه (درجه سانتی گراد) *Daily maximum temperature (°C)*	Tmax (°C)
دمای کمینه روزانه (درجه سانتی گراد) *Daily minimum temperature (°C)*	Tmin (°C)
بارندگی طی فصل رشد (میلی‌متر) *Precipitation during the growing season (mm)*	Prec (mm)
تبخیر و تعرق طی فصل رشد (میلی‌متر) *Evapotranspiration during the growing season (mm)*	ETGS (mm)
شاخص خشکی *Aridity index*	AI
تبخیر و تعرق (میلی‌متر) *Evapotranspiration (mm)*	ET (mm)
مجموع بارندگی در کل فصل رشد (میلی‌متر) *Total precipitation during the growing season (mm)*	PGS (mm)
تعداد روزهای بارندگی در فصل رشد *Number of rainy days during the growing season*	NPGS (روز)
دمای بیشینه در فصل رشد (درجه سانتی گراد) *Maximum temperature during the growing season (°C)*	TmaxGS (°C)
دمای کمینه در فصل رشد (درجه سانتی گراد) *Minimum temperature during the growing season (°C)*	TminGS (°C)
دمای میانگین در فصل رشد (درجه سانتی گراد) *Average temperature during the growing season (°C)*	TmeanGS (°C)
تعداد روزهای با دمای بیش از ۳۰ درجه سانتی گراد *Number of days with temperature above 30°C*	NDO30
طول فصل رشد (روز) *Length of the growing season (days)*	GSL

Climatic Variable	متغیر اقلیمی/ زراعی
TS	شدت فصلی بودن درجه حرارت *Temperature seasonality*
GDD	درجه روز رشد *Growing degree days (GDD)*
GDDGS (°C day)	درجه روز رشد طی فصل رشد (درجه سانتی گراد روز) *Growing degree days during the growing season (°C·days)*

طول فصل رشد، از ۱۵ آبان تا ۱۵ خرداد سال بعد در نظر گرفته شد.

این متغیرها در قالب یک دیتافریم منظم ترکیب شده و مراحل آماده‌سازی داده‌ها شامل نرمال‌سازی (بر اساس امتیاز زد)، کنترل کیفیت، و ادغام با داده‌های عملکرد روی آنها انجام گرفت.

تجزیه و تحلیل داده‌ها

۱ -پیش بینی عملکرد گندم به کمک الگوریتم های یادگیری ماشین

آموزش و تست مدل: در این مطالعه برای تحلیل اثر متغیرهای اقلیمی بر عملکرد گندم، از دو الگوریتم مستقل شامل رندوم فارست و ایکس‌جی‌بوست که هر دو از روش های یادگیری ماشین قدرتمند برای مدل‌سازی روابط پیچیده میان متغیرهای مستقل (ورودی‌ها) و متغیر وابسته (خروجی) محسوب می شوند در قالب سه رهیافت شامل پیش بینی کمی عملکرد گندم، پیش بینی طبقه‌ای عملکرد گندم، و پیش بینی عملکرد گندم در قالب ناحیه بندی اقلیمی استفاده شد. قابل یادآوری است که، از آنجایی که این دو الگوریتم نسبت به داده های استاندارد نشده، مقاوم هستند (بر خلاف شبکه های عصبی مصنوعی که هنگام کار با آنها باید داده ها استاندارد شوند)، بنابراین از داده های دارای مقیاس واقعی استفاده شد (به جز در مورد نشان دادن اهمیت نسبی ویژگی ها). قبل از اجرای فرآیند یادگیری، داده ها به دوقسمت تقسیم شدند، ۷۰ درصد داده ها جهت آموزش مدل و ۳۰ درصد برای تست مدل اختصاص یافتند.

تحلیل با رندوم فارست: این الگوریتم با ایجاد مجموعه‌ای از درختان تصمیم‌گیری و تلفیق نتایج آن‌ها از طریق رأی‌گیری یا میانگین‌گیری، پیش‌بینی نهایی را انجام می‌دهد. رگرسور اختصاصی این الگوریتم، رندوم فارست رگرسور نام دارد. ویژگی‌های برجسته رندوم فارست شامل موارد زیر است: توانایی کشف روابط پیچیده و غیرخطی میان متغیرهای اقلیمی و عملکرد محصول؛ مقاومت در برابر بیش‌برازش[1] به‌ویژه در حضور متغیرهای متعدد؛ ارائه رتبه‌بندی اهمیت متغیرها[2] که در تحلیل نقش عوامل اقلیمی بسیار ارزشمند است؛ دقت بالای پیش‌بینی حتی در صورت وجود داده‌های ناقص یا پُرنویز؛ کارایی در ساختارهای داده‌ای پنلی و سری‌زمانی بدون نیاز به فرض های محدودکننده.

به منظور ارزیابی عملکرد مدل رندوم فارست، از شاخص‌های آماری نظیرضریب تبیین[3] (R^2)، جذر میانگین مربعات خطا[4]، میانگین مطلق خطا[5] و شاخص توافق ویلموت[6] استفاده گردید. جهت تحلیل عمیق‌تر نتایج، مجموعه‌ای ازنمودارهای گرافیکی شامل تحلیل اهمیت نسبی متغیرها، درخت‌های تصمیم‌گیری همراه با شاخص های ارزیابی مدل برای هر شهرستان ترسیم شد. این نمودارها

[1] Overfitting
[2] Feature Importance
[3] Coefficient of determonation (R^2)
[4] Root Mean Square Error (RMSE)
[5] Mean Absolute Error (MAE)
[6] Willmott's Agreement Index (d)

درک دقیق‌تری از ترتیب و نقش متغیرهای اقلیمی و زراعی در تبیین عملکرد گندم فراهم کرده و به تفسیر روابط حاکم در هر شهرستان و به دنبال آن هر ناحیه اقلیمی کمک می‌کنند.

شایان ذکر است که برای هر شهرستان، یک مدل اختصاصی آموزش داده شد و پس از ارزیابی دقیق با داده‌های جدید، عملکرد آن با استفاده از شاخص‌های آماری تأیید شد. این مدل‌های آموزش‌دیده در قالب فایل‌هایی با فرمت pkl. ذخیره و مستندسازی گردیدند تا در کاربردهای آتی (جهت پیش بینی) مورد استفاده قرار گیرند.

تحلیل با ایکس‌جی‌بوست: کلیه مراحل انجام گرفته در الگوریتم رندوم فارست، با داده های ورودی یکسان، با الگوریتم ایکس‌جی‌بوست نیز انجام گرفتند. رگرسور اختصاصی این الگوریتم، ایکس‌جی‌بوست رگرسور نام دارد.

۲ -پیش بینی طبقه عملکرد گندم در شهرستان های استان خراسان رضوی

به منظور طبقه بندی (کلاس بندی) عملکرد گندم در هر شهرستان، از دو الگوریتم رندوم فارست و ایکس‌جی‌بوست استفاده شد، با این تفاوت که در این رهیافت، مدل ها بهینه سازی نشدند (به دلیل ماهیت طبقه ای داده ها در این رویکرد).

بهینه سازی مدل ها: جهت بهینه‌سازی ابرپارامترهای مدل‌های یادگیری ماشین و یافتن بهترین ترکیب ممکن برای هر الگوریتم، از دو رویکرد جستجوی شبکه‌ای با اعتبارسنجی متقابل[1] و جستجوی تصادفی با اعتبارسنجی متقابل[2] بهره‌گیری شد(تعداد فولد، ۵ و n_estimator برابر با ۳۰۰). رویکرد اول تمام ترکیب‌های ممکن از پارامترهای الگوریتم را (مثلاً عمق درخت، تعداد درخت، نرخ یادگیری و غیره) را تولید کرده و آنها را تست می‌کند و در ادامه بهترین آنها را بر اساس دقت (مثلاً R^2) برای متغیر هدف (مثلاً عملکرد گندم) پیدا می‌کند، در حالی‌که رویکرد دوم با انتخاب تصادفی اما هدفمند، امکان بهینه‌سازی سریع‌تر را در مسائل با فضای پارامتر گسترده فراهم می‌سازد. تنها عاملی که به کارگیری رویکرد اول را می تواند محدود کند، این است که زمانی که تعداد پارامترها زیاد باشد، سرعت اجرای آن روی رایانه های ضعیف، کُند خواهد بود. در این پژوهش، مدل‌های رندوم فارست و ایکس‌جی‌بوست با استفاده از این دو روش، بهینه‌سازی و با اعتبارسنجی متقابل پنج‌تایی[3] تعیین اعتبار شدند. در جدول ۳ این دو رویکرد با یکدیگر مقایسه شده اند.

جدول ۳- مقایسه دو رویکرد بهینه سازی الگوریتم های یادگیری ماشین در پیش بینی عملکرد گندم

Table 3 - Comparison of Two Optimization Approaches for Machine Learning Algorithms in Wheat Yield Prediction

ویژگی/الگوریتم Feature / Algorithm	جستجوی تصادفی Randomized Search	جستجوی شبکه‌ای Grid Search
دقت جستجو Search Accuracy	متوسط، (تعدای ترکیب تصادفی) Moderate (a number of random combinations)	بالا، (تمام ترکیب‌های ممکن) High (all possible combinations)
سرعت Speed	سریع تر Faster	کند، مخصوصاً با پارامتر زیاد Slower, especially with many parameters
استفاده بهینه از منابع Efficient Use of Resources	بله Yes	نه زیاد Not much
برای تست پارامترهای زیاد در زمان محدود For testing many parameters in limited time	خیلی خوب Very good	کند Slow
برای تست پارامترهای کم در زمان کافی	خیلی خوب Very good	عالی Excellent

[1] GridSearch Cross Validation
[2] RandomizedSearch Cross Validation
[3] 5-Fold Coss-Validation

تعیین اعتبار[1] و ارزیابی دقت مدل[2] ها: برای ارزیابی عملکرد مدل‌های یادگیری ماشین، فرایند تعیین اعتبار یا اعتبارسنجی نقش کلیدی ایفا می‌کند. در این فرایند، مدل‌ها روی داده‌هایی که در آموزش شرکت نداشته‌اند مورد آزمایش قرار می‌گیرند تا میزان تعمیم‌پذیری آن‌ها بررسی شود. در فرآیند مدل‌سازی عملکرد گندم، تعیین اعتبار نقشی حیاتی در اطمینان از قابلیت تعمیم مدل به داده‌های جدید ایفا می‌کند. یکی از مهم‌ترین جنبه‌های اعتبارسنجی، ارزیابی دقت است که با استفاده از معیارهایی شامل ضریب تبیین، جذر میانگین مربعات خطا، میانگین مطلق خطا، و شاخص توافق ویلموت (دی)، کیفیت پیش‌بینی‌های مدل را کمّی می‌سازد. بنابراین، ارزیابی دقت را می‌توان بخشی از فرایند جامع‌تر تعیین اعتبار دانست که امکان مقایسه‌ی دقیق بین مدل‌ها و انتخاب گزینه‌ی بهینه را فراهم می‌سازد. بدین ترتیب، دقت مدل‌ها برای هر ناحیه اقلیمی به‌صورت مجزا سنجیده شد تا میزان اطمینان‌پذیری آن‌ها در مناطق با شرایط متفاوت اقلیمی بررسی گردد. این تحلیل، مبنایی برای انتخاب مدل نهایی و تفسیر نتایج نهایی پیش‌بینی فراهم ساخت.

برای ارزیابی دقت مدل‌های طبقه‌بندی، از ماتریس درهم‌ریختگی[3] استفاده شد که امکان تحلیل دقیق عملکرد مدل را فراهم می‌سازد. این ماتریس، تعداد پیش‌بینی‌های درست و نادرست را برای هر کلاس به‌صورت مجزا نشان می‌دهد و شامل چهار مقدار کلیدی است: مثبت واقعی[4]، منفی واقعی[5]، مثبت نادرست[6]، و منفی نادرست[7]. با استفاده از این مقادیر، معیارهای مهمی نظیر دقت کلی یا صحت[8]، بازیابی یا حساسیت[9]، امتیاز اف- وان[10] و دقت مثبت پیش‌بینی[11] محاسبه می‌شوند (Ting, 2011). این تحلیل کمک می‌کند تا نقاط ضعف مدل در شناسایی کلاس‌ها به‌ویژه در داده‌های نامتوازن مشخص شود. در این مطالعه، از ماتریس درهم ریختگی برای مقایسه عملکرد الگوریتم های رندوم فارست و ایکس‌جی‌بوست در دسته‌بندی شهرستان های با سه کلاس عملکرد بالا، متوسط و پایین استفاده شد. برای این منظور مراحل زیر انجام گرفت:

۱- جمع زدن پیش‌بینی‌ها درون هر خوشه. ۲- ایجاد بردار برای نمونه واقعی و پیش‌بینی‌شده (y_true_clusterN شامل برچسب‌های واقعی عملکرد مثلاً high/low و y_pred_clusterN برای پیش‌بینی‌های مدل. ۳- ایجاد کانفیوژن ماتریکس برای هر خوشه.

تحلیل و ارزیابی نقش متغیرها بر اساس چارچوب SHAP

به‌منظور تبیین نقش نسبی هر یک از ویژگی‌های ورودی در پیش‌بینی عملکرد گندم، از چارچوب تفسیری SHapley Additive exPlanations (SHAP) استفاده شد. روش SHAP یکی از قدرتمندترین ابزارهای تفسیر مدل‌های یادگیری ماشین است که با بهره‌گیری از مفاهیم نظریه بازی شاپلی، سهم مشارکتی هر ویژگی را در خروجی مدل به‌صورت منصفانه و دقیق برآورد می‌کند. این رویکرد، برخلاف برخی روش‌های سنتی که صرفاً به ارزیابی حضور یا عدم حضور یک متغیر بسنده می‌کنند،

[1] Validation
[2] Model Evaluation
[3] Confusion Matrix
[4] True Positive (TP)
[5] True Negative (TN)
[6] False Positive (FP)
[7] False Negative (FN)
[8] Accuracy
[9] Recall
[10] F1-Score
[11] Precision

قادر است تأثیر هر متغیر را در بسترهای مختلف داده و در سطح محلی[1] (برای یک نمونه خاص) و سراسری[2] (میانگین مقادیر شیپ برای همه نمونه‌ها به‌دست می‌آید و اهمیت کلی ویژگی‌ها مشخص می‌شود) مورد تحلیل قرار دهد. در نتیجه، تحلیل مبتنی بر شیپ تصویری جامع از نحوه و شدت تأثیرگذاری هر متغیر در سراسر فضای داده‌ها ارائه می‌دهد (Remman and Lekkas, 2021; Kim and Kim, 2022). مهمترین دلایل بهتر بودن شیپ‌سایر روش های مشابه به قرار زیر است: سازگار بودن با هر نوع مدل مانند مدل های دیپ، درختی، و خطی. نشان دادن سهم منصفانه هر ویژگی (طبق قضایای ریاضی)؛ بر خلاف روش‌هایی مثل اهمیت بر پایه جابجایی[3]، اثر تعاملی و همبستگی بین ویژگی‌ها را نیز در نظر می‌گیرد. در حالی که اهمیت بر پایه جابجایی با جابجا کردن مقادیر هر ویژگی در داده‌های آزمون و سنجش تغییر در عملکرد مدل، میزان اهمیت هر ویژگی را برآورد می‌کند، روش شیپ بر پایه نظریه بازی‌های مشارکتی، سهم منصفانه و دقیق هر ویژگی را در پیش‌بینی خروجی مدل – هم به صورت محلی برای هر نمونه و هم به صورت کلی – اندازه‌گیری می‌کند. بر خلاف اهمیت بر پایه جابجایی که ممکن است در حضور ویژگی‌های همبسته دچار خطا یا تفسیر گمراه‌کننده شود، شیپ توانایی جداسازی و تحلیل دقیق اثرات فردی و تعاملی ویژگی‌ها را دارد. به همین دلیل، در تحلیل‌های مبتنی بر مدل‌های پیچیده مانندرندوم فارست و ایکس جی بوست، استفاده از رهیافت شیپ توصیه می‌شود. برای نمایش نتایج این تحلیل، از نمودارهای ویولن[4] استفاده شد. این نمودارها توزیع مقادیر شیپ مرتبط با هر ویژگی را نشان می‌دهند و از طریق شکل ترکیبی خود (شامل چگالی احتمال و آمار توصیفی)، هم جهت اثرگذاری (مثبت یا منفی بودن شیپ به این معنی که ویژگی‌های با مقادیر شیپ باعث افزایش مقدار پیش‌بینی شده می‌شوند. ویژگی‌های با مقادیر شیپ منفی باعث کاهش مقدار پیش‌بینی شده می‌شوند) و هم شدت نسبی آن را بر خروجی مدل به تصویر می‌کشند.

نمودار دیگری که شیپ تولید می کند، نمودار تصمیم[5] نام دارد که نشان می دهد چگونه هر ویژگی بر پیش بینی مقدار متغیر وابسته اثر می گذارد. آخرین نموداری که به کمک شیپ برای هر شهرستان ایجاد شد، نمودار وابستگی جزئی[6] است که رابطهٔ مستقیم بین هر ویژگی و متغیر وابسته (عملکرد گندم در هر شهرستان) را به تصویر می کشد. استفاده از این نمودارها، ضمن تقویت شفافیت مدل‌های یادگیری ماشین پیچیده، زمینه را برای تفسیر علمی خروجی مدل و شناسایی عوامل کلیدی مؤثر بر عملکرد گندم فراهم می‌سازد؛ موضوعی که برای تدوین راهبردهای مدیریتی و سیاست‌گذاری کشاورزی از اهمیت بالایی برخوردار است.

۳- پیش بینی عملکرد گندم بر اساس خوشه بندی شهرستان ها در ناحیه های اقلیمی: برای شناسایی نواحی اقلیمی-کشاورزی مشابه (همگن) در استان خراسان رضوی، از روش خوشه‌بندی سلسله‌مراتبی[7] استفاده شد. این روش برای تحلیل سه شاخص اقلیمی-کشاورزی شامل شاخص خشکی، شدت فصلی بودن درجه حرارت و درجه روز رشد به‌کار گرفته شده است. خوشه‌بندی سلسله‌مراتبی به‌دلیل توانایی آن در شناسایی الگوهای پیچیده و توزیع نواحی با ویژگی‌های مشابه، انتخابی مناسب برای تحلیل داده‌های توزیع شده از چندین شاخص اقلیمی-کشاورزی است. این روش به‌ویژه در این حالت که داده‌ها از چندین شاخص مختلف استخراج شده‌اند، دقت بالاتری در شناسایی نواحی با ویژگی‌های مشابه ارائه می‌دهد. در پژوهش حاضر، ابتدا شاخص‌های

[1] local
[2] global
[3] permutation importance
[4] Violin plots
[5] Decision Plot
[6] Partial Dependence Plot
[7] Hierarchical Clustering

اقلیمی-کشاورزی در محیط GIS درون‌یابی شدند و سپس مجموعه‌ای از نقاط نمونه‌ای (تصادفی) در سراسر منطقه استخراج شدند (۳۵۰ نقطه). این نقاط به‌عنوان ورودی تحلیل خوشه‌بندی در نرم‌افزار R و با استفاده از بسته‌های factoextra و cluster پردازش شدند. مراحل تحلیل به‌ترتیب زیر انجام شد:

- پیش‌پردازش داده‌ها با حذف مقادیر نامعتبر و استانداردسازی (scale) شاخص‌های ورودی.
- محاسبه ماتریس فاصله با استفاده از روش فاصله اقلیدسی[1] بین نقاط نمونه‌ای.
- اجرای خوشه‌بندی سلسله‌مراتبی و گروه‌بندی نقاط مشابه در خوشه‌های مجزا با استفاده از روش نسخه بهبودیافته از روش Ward (Ward.D2) در بسته cluster.
- انتخاب تعداد بهینه خوشه‌ها (k) با استفاده از شاخص Silhouette. با افزایش آستانه شباهت (که باعث ایجاد خوشه‌های بزرگ‌تر و عمومی‌تر می‌شود)، شاخص سیلوهت کاهش می یابد. با گرفتن خروجی های متعدد، در نهایت بر اساس شاخص سیلوهت ۰/۴۱ (برابر با نزدیکی[2] یا شباهت[3] ۶۰ درصد) خوشه بندی انجام گرفت. مقدار ۰/۴۱ بهترن توازن بین تفکیک و همگنی خوشه ها را ایجاد کرد.
- انتساب خوشه‌ها به داده‌ها (برچسب‌گذاری) بر اساس نتایج خوشه‌بندی، به هر نقطه نمونه‌ای.

در نهایت، نتایج حاصل از خوشه‌بندی به فرمت CSV ذخیره شده و برای تحلیل فضایی به محیط ArcGIS منتقل شدند. در این مرحله، داده‌های نقطه‌ای خوشه‌بندی‌شده با استفاده از روش کریجینگ ساده درون‌یابی شده و به یک سطح پیوسته تبدیل شدند. سپس، با طبقه‌بندی نقشه‌ی درون‌یابی‌شده، محدوده‌های اقلیمی متناظر با هر خوشه به‌صورت پلی‌گون[4] استخراج شدند. این خروجی‌های پلی‌گونی مبنای تحلیل‌های بعدی در مطالعه حاضر قرار گرفته و نواحی اقلیمی-کشاورزی استان را شناسایی کرده‌اند. توصیف هر ناحیه اقلیمی مطابق طبقه‌بندی اقلیمی شهرستان‌ها بر اساس سیستم کوپن–گایگر و آمار بروز شده سایت کلایمیت دیتا صورت گرفت(Peel et al., 2007; http://en.climate-data.org/asia/iran/razavi-khorasan-2216/) . همچنین، به منظور یادگیری بهتر مدل ها، شماره ناحیه اقلیمی هر شهرستان در یک ستون مستقل قرار گرفت. مزیت افزودن این متغیر در کنار سایر متغیرها این است که مدل، تفاوت اقلیم‌ها رو یاد می‌گیرد، ولی همچنان از تمام جزئیات در سطح شهرستان بهره می‌برد، به عبارت دیگر، تفکیک پذیری مدل بهبود می یابد.

نرم افزارهای مورد استفاده

در این پژوهش، الگوریتم های یادگیری ماشین، رندوم فارست و ایکس‌جی‌بوست، با استفاده از زبان برنامه‌نویسی پایتون Python 3.3 و کتابخانه Scikit-learn پیاده‌سازی شدند. آماده سازی دیتافریم ها و خوشه بندی شهرستان های استان در قالب ناحیه های اقلیمی، با استفاده از زبان برنامه نویسی R 4.3.2 و نرم افزار ArcGIS 10.8.2 انجام گرفت. از آنجایی که هر دو الگوریتم مورد استفاده در این پژوهش درخت‌محور هستند، و درخت‌ها به مقیاس داده حساس نیستند، و می توانند با داده‌های خام نیز کار کنند بنابراین، نیازی به نرمال‌سازی یا استانداردسازی داده ها نیست. برای اجرای رندوم فارست در مواردی که مقیاس داده ها خیلی متفاوت باشد (مثلاً اگر بازه داده های تبخیر و تعرق بین ۰ تا ۵ و درجه روز رشد بین ۱۰۰۰ تا ۳۰۰۰ باشد)، پیشنهاد شده در صورت

[1] Euclidean Distance
[2] distance
[3] similarity
[4] Polygon

امکان داده ها استاندارد شوند، اگر چه که اجباری در این کار نیست. در یک مورد خاص (نمودارهای شیپ ویولونی)، داده ها جهت استفاده از ابزار شیپ استاندارد و خروجی ها بر این مبنا رسم شدند.

نتایج و بحث

تحلیل، مقایسه و ارزیابی مدل های پیش بینی عملکرد گندم آبی در سطح استان خراسان رضوی، طی دو رهیافت مستقل انجام گرفت. در رهیافت اول، دو مدل رندوم فارست و ایکس‌جی‌بوست، آموزش داده و بهینه شدند، سپس روی داده هایی که درآموزش شرکت نداشتند، تست شدند. در رهیافت دوم، همین دو مدل، داده های عملکرد گندم را کلاس بندی کرده و سپس عملکرد دو مدل بر روی نتایج تحلیل شد. در جدول ۴ معیارهای ارزیابی دو مدل در رهیافت اول، قبل و بعد از بهینه سازی نشان داده شده اند.

جدول ٤- معیارهای ارزیابی و تعیین اعتبار دو الگوریتم رندوم فارست و ایکس‌جی‌بوست در پیش بین عملکرد گندم آبی در استان خراسان رضوی، قبل و بعد از بهینه سازی

Table 4 - Evaluation and Validation Metrics for Random Forest and XGBoost Algorithms in Predicting Irrigated Wheat Yield in Khorasan Razavi Province, Before and After Optimization

ModelStatus/Evaluation Metric معیار ارزیابی/ وضعیت مدل	RMSE	R^2	MAE	d
RF (Befor Optimization)(رندوم فارست (پیش از بهینه سازی	399.53	0.134	326.24	0.46
RF(After Optimization)(رندوم فارست (پس از بهینه سازی	395.13	0.136	331.50	0.47
XGB (Befor Optimization)(ایکس‌جی‌بوست (پیش از بهینه سازی	499.93	0.063	413.02	0.46
XGB (After Optimization)(ایکس‌جی‌بوست (پس از بهینه سازی	492.21	0.074	408.09	0.42

بهینه سازی مدل رندوم فارست، جذر میانگین مربعات خطا را به مقدار چهار واحد کاهش داد، و ضریب همبستگی را به مقدار دو هزارم افزایش داد. بهینه سازی مدل ایکس‌جی‌بوست، تمام معیارهای ارزیابی (به جز شاخص توافق ویلموت) را نسبت به مدل بهینه نشده، بهبود داد. بنابراین، در ادامه نتایج مربوط به الگوریتم های بهینه سازی شده (به غیر از رویکرد طبقه بندی) آورده می شوند.

تحلیل ویژگی‌های مؤثر در دو الگوریتم برای شهرستان ها

برای درک بهتر از رفتار مدل‌های یادگیری ماشین و بررسی دلایل پیش‌بینی‌های انجام‌شده، نمودارهای شیپ یا ویولونی اهمیت ویژگی‌ها (فیچر ایمپورتنس) (شکل ۱ آ و ب)، نمودارهای ضرایب همبستگی بین ده ویژگی برتر (شکل ۳ آ و ب) برای هر شهرستان به تفکیک و برای هر دو مدل رندوم‌فارست و ایکس‌جی‌بوست تولید شدند.

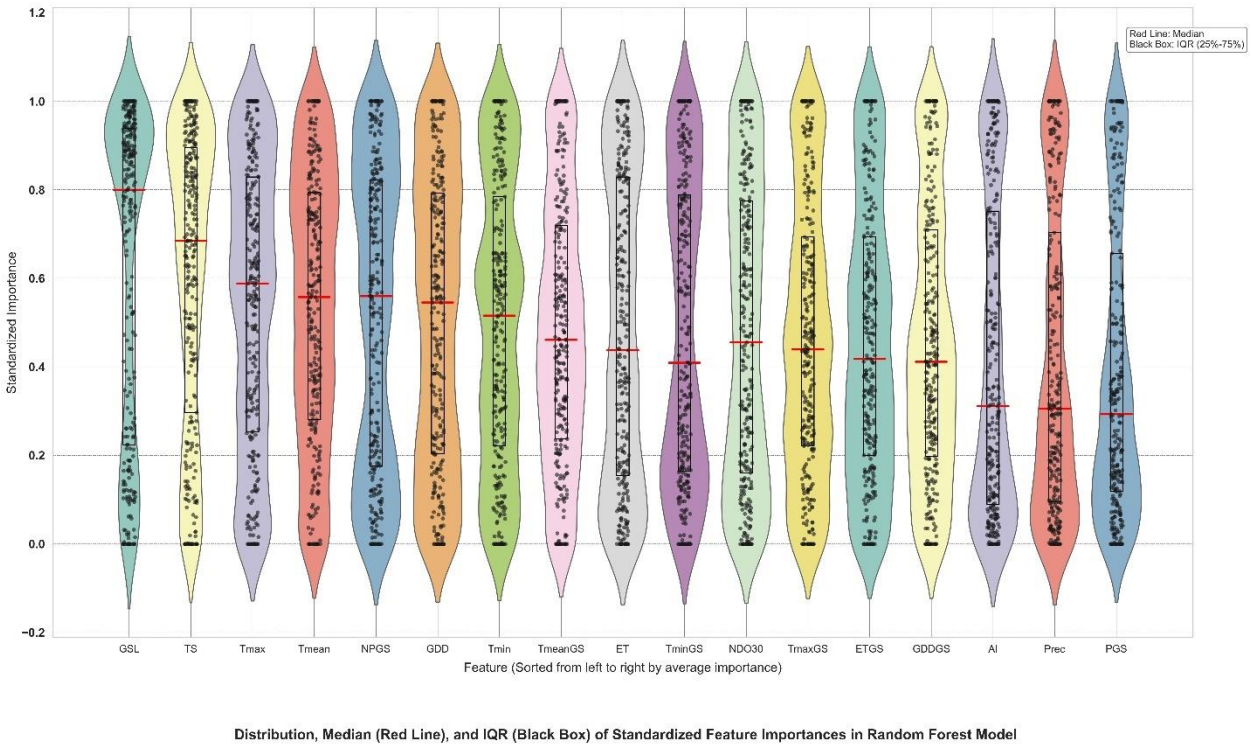
توزیع مقادیر شیپ (اهمیت نسبی ویژگی‌ها) در مدل رندوم فارست

شکل ۱، توزیع اهمیت استانداردشده ویژگی‌های اقلیمی و فیزیولوژیکی مؤثر بر عملکرد گندم را بر اساس مدل رندم فارست نشان می‌دهد. ویژگی‌ها بر اساس میانگین اهمیت از چپ به راست مرتب شده‌اند و هر نمودار ویولن، تراکم و پراکندگی مقادیر اهمیت را برای یک ویژگی خاص به تصویر می‌کشد. خط قرمز نشان‌دهنده میانه[1] و جعبه سیاه بیانگر فاصله بین چارک اول و سوم[2] است. نقاط سیاه رنگ داخل ویولون ها، نشان‌دهنده‌ی مقدار اهمیت هر ویژگی در نمونه‌های مختلف (مثلاً در تکرارهای cross-validationیا در بین زیربخش‌های داده هستند). وقتی این نقاط در یک متغیر خاص، به‌جای تمرکز در یک محدوده‌ی باریک،

[1] Median
[2] IQR

در کل محور y پخش شـده‌اند، نشـان می‌دهد که اهمیت آن متغیر در شـرایط مختلف (یا نمونه‌های مختلف) مدل، نوسـان زیادی دارد، به عبارت دیگر، مدل نسـبت به این ویژگی در برخی موارد آن را خیلی مهم و در موارد دیگر کم‌اهمیت ارزیابی کرده، که می‌تواند ناشی از شرایط اقلیمی متغیر، اثرات فصلی، یا تعاملات پیچیده با سایر متغیرها باشد.

متغیرهایی مانند طول فصل رشد، شدت فصلی بودن درجه حرارت، و دمای بیشینه روزانه در سمت چپ نمودار قرار دارند و دارای بیشـترین میزان اهمیت میانگین هسـتند. پهنای زیاد در ناحیه بالایی این ویولن‌ها نشـان می‌دهد که مقدار اهمیت این ویژگی‌ها در بین نمونه‌های مختلف به‌طور گسترده در سطح بالایی متمرکز است. در مقابل، ویژگی‌هایی مانند شاخص خشکی، بارندگی، و بارندگی طی فصـل رشـد با تراکم در سـطوح پایین‌تر و میانه پایین‌تر، نقش کم‌تری در پیش‌بینی عملکرد مدل داشـته‌اند. این نشـان می‌دهد که در اغلب تکرارهای مدل، این ویژگی‌ها تأثیر نسـبتاً کم‌تری در خروجی نهایی داشـته‌اند. پراکندگی نقاط سـیاه نشـان‌دهنده نوسـان اهمیت در بین نمونه‌هاسـت. وجود نقاط با پراکندگی عمودی زیاد مثل تعداد دفعات وقوع بارندگی طی فصـل رشـد و دمای کمینه طی فصل رشد، نشان‌دهنده وابستگی متغیر به موقعیت مکانی یا شرایط خاص داده‌ها است، در حالی که متغیرهایی با توزیع فشرده‌تر مثل طول فصل رشد و دمای بیشینه طی فصل رشد نشان‌دهنده ثبات بالاتر در اهمیت هستند.



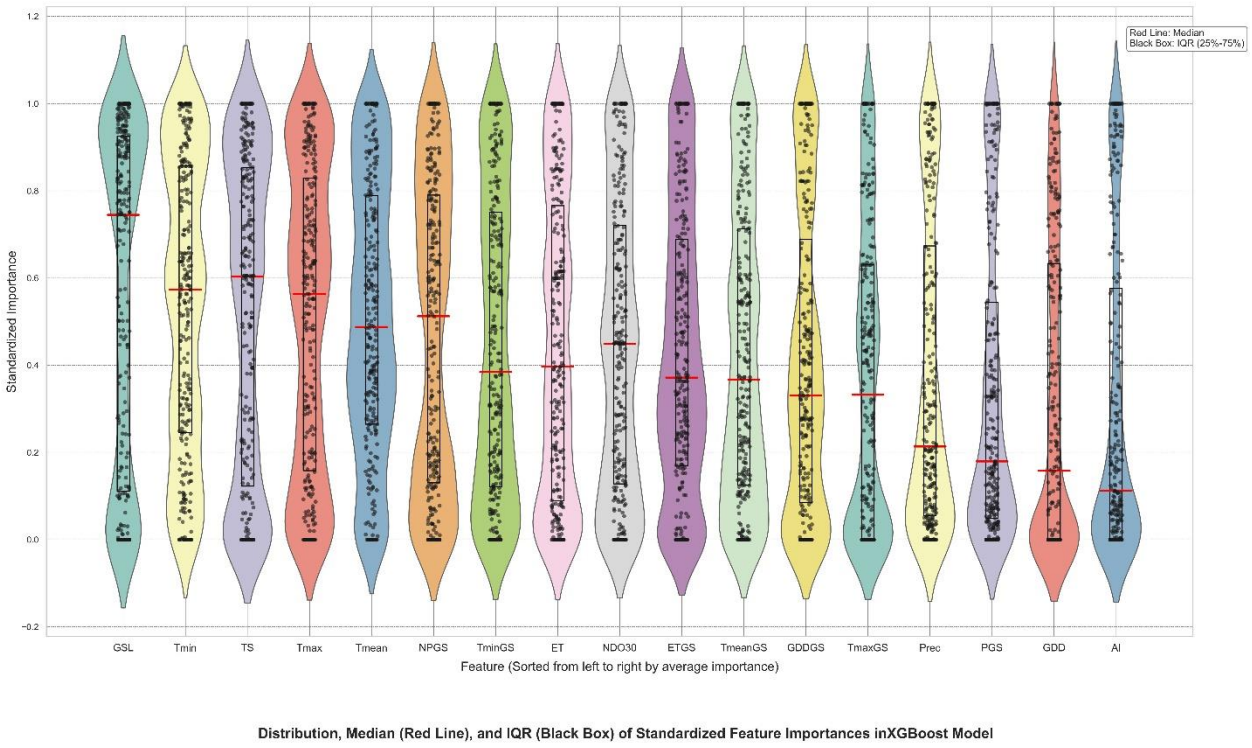
Distribution, Median (Red Line), and IQR (Black Box) of Standardized Feature Importances in Random Forest Model

شکل ۱- نمودارهای ویولونی توزیع مقادیر شیپ (اهمیت ویژگی‌ها) در الگوریتم رندوم فارست جهت پیش بینی عملکرد گندم در استان خراسان رضوی. (شرح کامل ویژگی ها در جدول ۲ بخش مواد و روش ها آورده شده است).

Figure 1 – Violin plots of SHAP value distributions (feature importance) in the Random Forest algorithm for predicting wheat yield in Khorasan Razavi province. (A complete description of the features is provided in Table 2 in the Materials and Methods section).

توزیع مقادیر شیپ (اهمیت نسبی ویژگی‌ها) در مدل ایکس‌جی‌بوست

شکل ۲ توزیع اهمیت استانداردشده متغیرها را در مدل ایکس جی بوست نمایش می‌دهد. در این نمودار نیز مانند شکل قبلی، خط قرمز نشان‌دهنده میانه و جعبه سیاه نشان‌دهنده فاصله بین چارک اول و سوم است. ویژگی‌ها بر اساس میانگین اهمیت مرتب شده‌اند. متغیرهایی مانند طول فصل رشد، دمای کمینه روزانه، شدت فصلی بودن درجه حرارت، و دمای بیشینه روزانه در سمت چپ نمودار و دارای بیشترین اهمیت میانگین هستند، به‌ویژه متغیر طول فصل رشد مجدداً بالاترین سطح اهمیت را نشان می‌دهد، با توزیع متراکم در نواحی بالای محور، که نشان از تأثیر پایدار آن در بین نمونه‌های مختلف دارد. متغیرهایی نظیر شاخص خشکی، درجه روز رشد، و بارندگی طی فصل رشد در سمت راست قرار دارند و با تمرکز داده‌ها در بخش‌های پایین‌تر، از نظر مدل اهمیت نسبتاً کمی دارند. در برخی متغیرها مانند دمای کمینه طی فصل رشد، تبخیر و تعرق یا تعداد روزهای با دمای بالای ۳۰ درجه سانتی گراد، پراکنش عمودی قابل توجه نقاط درون نمودار ویولن، بیانگر نوسانات بالای اهمیت این ویژگی‌ها در بین نمونه‌هاست. این امر می‌تواند حاکی از حساسیت مدل نسبت به شرایط متغیر محیطی یا تعاملات زمانی آن‌ها باشد. مقادیر میانه برای بسیاری از متغیرهایی که در ناحیه وسط نمودار واقع شده اند، مثل تبخیر و تعرق طی فصل رشد، دمای میانگین طی فصل رشد یا درجه روز رشد طی فصل رشد نزدیک به ۰/۵ تا ۰/۶ قرار دارد که نشان از نقش میان‌رده آنها در تصمیمات مدل دارد. تحلیل توزیع مقادیر شیپ، می‌تواند به شناسایی عوامل کلیدی تأثیرگذار بر عملکرد گندم در مدل‌های یادگیری ماشین کمک کرده و مبنایی برای تصمیم‌گیری در برنامه‌ریزی‌های زراعی فراهم آورد.



Distribution, Median (Red Line), and IQR (Black Box) of Standardized Feature Importances inXGBoost Model

شکل ۲- نمودارهای شیپ یا ویولونی اهمیت ویژگی‌های بکار گرفته شده در مدل ایکس جی بوست جهت پیش بینی عملکرد گندم در استان خراسان رضوی. ((شرح کامل ویژگی ها در جدول ۲ بخش مواد و روش ها آورده شده است).

Figure 2 – Violin plots of SHAP value distributions (feature importance) in the XGBoost algorithm for predicting wheat yield in Khorasan Razavi province. (A complete description of the features is provided in Table 2 in the Materials and Methods section).

مقایسه مدل رندوم فارست و ایکس جی بوست در ارزیابی اهمیت ویژگی‌ها

- در هر دو مدل، متغیر (طول فصل رشد) بیشترین اهمیت را دارد، که نشان‌دهنده‌ی تأثیر کلیدی این متغیر در پیش‌بینی عملکرد گندم است.
- در مدل رندوم فارست، متغیر دمای بیشینه روزانه نیز در میان متغیرهای بسیار مهم قرار داشت، اما در مدل ایکس جی بوست به جایگاه پایین‌تری سقوط کرده است، در حالی که دمای کمینه روزانه نقش مهم‌تری در مدل ایکس جی بوست ایفا می‌کند.
- مدل ایکس جی بوست تنوع بیشتری در پراکندگی اهمیت ویژگی‌ها نشان می‌دهد، به ویژه در متغیرهای با اهمیت میانه، که می‌تواند نشانگر قدرت تفکیک دقیق‌تر این مدل در شرایط پیچیده‌تر باشد.
- مقایسه نمودارها نشان می‌دهد که مدل رندوم فارست اهمیت نسبتاً متوازنی برای متغیرها در نظر گرفته، در حالی که مدل ایکس جی بوست تفاوت‌های بیشتری در اهمیت ویژگی‌ها قائل شده و نسبت به برخی متغیرهای خاص، حساسیت و تمایز بالاتری نشان داده است، به عبارت دیگر، تفکیک دقیق‌تری بین متغیرهای با اهمیت بالا و پایین ایجاد کرده است.

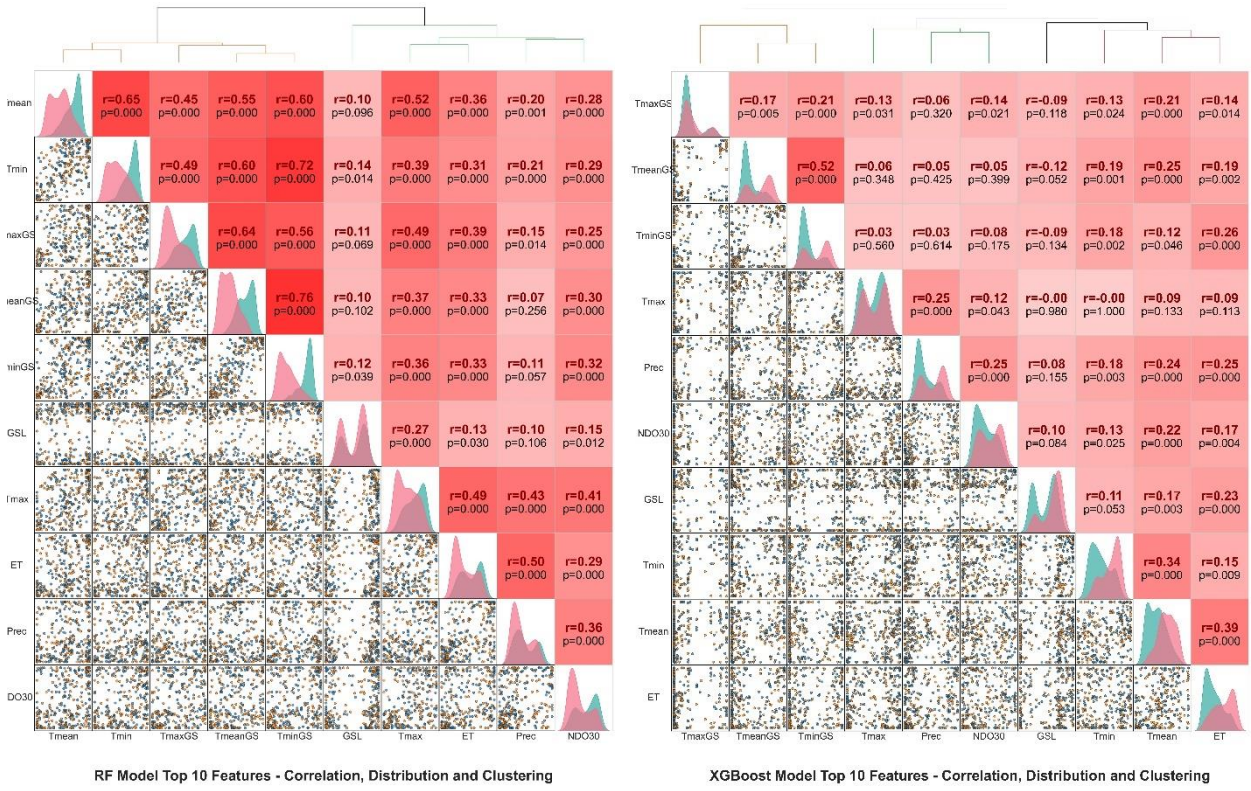
اهمیت هر ویژگی در مدل نشان می‌دهد که مدل چقدر به آن ویژگی برای پیش‌بینی‌ها اعتماد کرده است. به عبارت دیگر، این ویژگی‌ها تأثیر زیادی بر پیش‌بینی‌ها دارند و بر توان و دقت مدل‌های مختلف تأثیرگذار هستند. برای مثال، ویژگی شدت فصلی بودن درجه حرارت و دمای کمینه روزانه، دمای میانگین روزانه و طول فصل رشد اهمیت و تأثیر زیادی در پیش‌بینی‌های کلیه شهرستان ها دارد، در حالی که ویژگی ای مانند شاخص خشکسالی اهمیت و تأثیر کمتری دارند (در هر دو مدل). به نظر می رسد که ویژگی شدت فصلی بودن در جه حرارت، شاید به دلیل ارتباط قوی‌ای که با شرایط آب و هوایی یا ویژگی‌های محیطی دارد، تأثیر بیشتری در پیش‌بینی‌ها داشته باشد. از سوی دیگر، اهمیت کمتر ویژگی شاخص خشکسالی، احتمالاً به علت نزدیک بودن مقدار عددی آن برای تمام شهرستان ها به دلیلی واقع شدن در گستره ای بزرگ تر از یک منطقه خشک و نیمه خشک است. به کمک نمودارهای شیپ (اهمیت ویژگی ها) می توان نسبت به شناسایی و تعیین اولویت متغیرهای اقلیمی و زراعی و در ادامه برنامه ریزی و سیاست گذاری اقدام نمود.

برآورد تبخیر و تعرق مرجع روزانه با استفاده از روش جنگل تصادفی بهینه شده با الگوریتم ژنتیک در منطقه آذربایجان شرقی نشان داد که الگوریتم جنگل تصادفی نسبت به روش های تجربی در ایستگاه های مورد مطالعه، عملکرد بهتری داشت (Monavar Sabegh et al., 2023).

ضریب همبستگی بین ویژگی های با اهمیت نسبی بالا در دو مدل

شکل ۳ ماتریس همبستگی ده ویژگی برتر انتخاب‌شده توسط الگوریتم رندوم فارست را برای تمام شهرستان‌ها نشان می‌دهد و با بهره‌گیری از خوشه‌بندی سلسله‌مراتبی، ویژگی‌های مشابه را در کنار هم قرار داده است. این شکل نشان می‌دهد که بین دمای بیشینه طی فصل رشد، دمای میانگین طی فصل رشد، دمای میانگین روزانه، و دمای کمینه روزانه همبستگی‌های مثبت و نسبتاً قوی وجود دارد. به‌ویژه، بیشترین ضریب همبستگی بین دمای میانگین طی فصل رشد و دمای کمینه طی فصل رشد مشاهده شد ($r = 0.76$, $p < 0.001$)، و پس از آن، رابطه‌ی بین دمای کمینه روزانه و دمای کمینه فصل رشد با ضریب $r = 0.72$ قرار دارد. خوشه‌بندی ویژگی‌ها نشان می‌دهد دمای میانگین طی فصل رشد، دمای بیشینه طی فصل رشد و دمای کمینه طی فصل رشد در یک خوشه، و دما کمینه روزانه و دمای میانگین روزانه در خوشه‌ای مجزا قرار گرفته‌اند، که بیانگر تفاوت در الگوهای هم‌تغییری آن‌ها است.

Iranian Journal of
Field Crops Research



شکل ۳- ضرایب همبستگی، پی ولیو و توزیع چگالی و خوشه بندی ده ویژگی برتر مؤثر بر عملکرد گندم در استان خراسان رضوی در دو الگوریتم رندوم فارست (سمت چپ) و ایکس جی بسوت (سمت راست). (شرح کامل ویژگی ها در جدول ۲ بخش مواد و روش ها آورده شده است).

Figure 3 – Correlation coefficients, p-values, density distributions, and clustering of the top ten most influential features on wheat yield in Khorasan Razavi province using the Random Forest algorithm (left) and XGBoost algorithm (right). (A complete description of the features is provided in Table 2 in the Materials and Methods section).

در مدل ایکس جی بوست، بالاترین ضریب همبستگی میان دو ویژگی دمای کمینه طی فصل رشد و دمای میانگین طی فصل رشد مشاهده شد($r = 0.52, p < 0.01$) . این رابطه پیش‌تر در مدل رندوم فارست نیز به‌دست آمده بود، اما با شدت بیشتری ($r = 0.72, p < 0.01$). در میان سایر متغیرهای دمایی در مدل ایکس جی بوست، همبستگی مثبت و نسبتاً قوی بین دمای میانگین روزانه و دمای کمینه روزانه دیده شد($r = 0.34, p < 0.01$) . دومین ضریب همبستگی قابل‌توجه مربوط به رابطه بین تبخیر و تعرق طی فصل رشد و دمای میانگین روزانه بود($r = 0.39, p < 0.01$) ، که از نظر فیزیولوژیکی نیز قابل پیش‌بینی است.

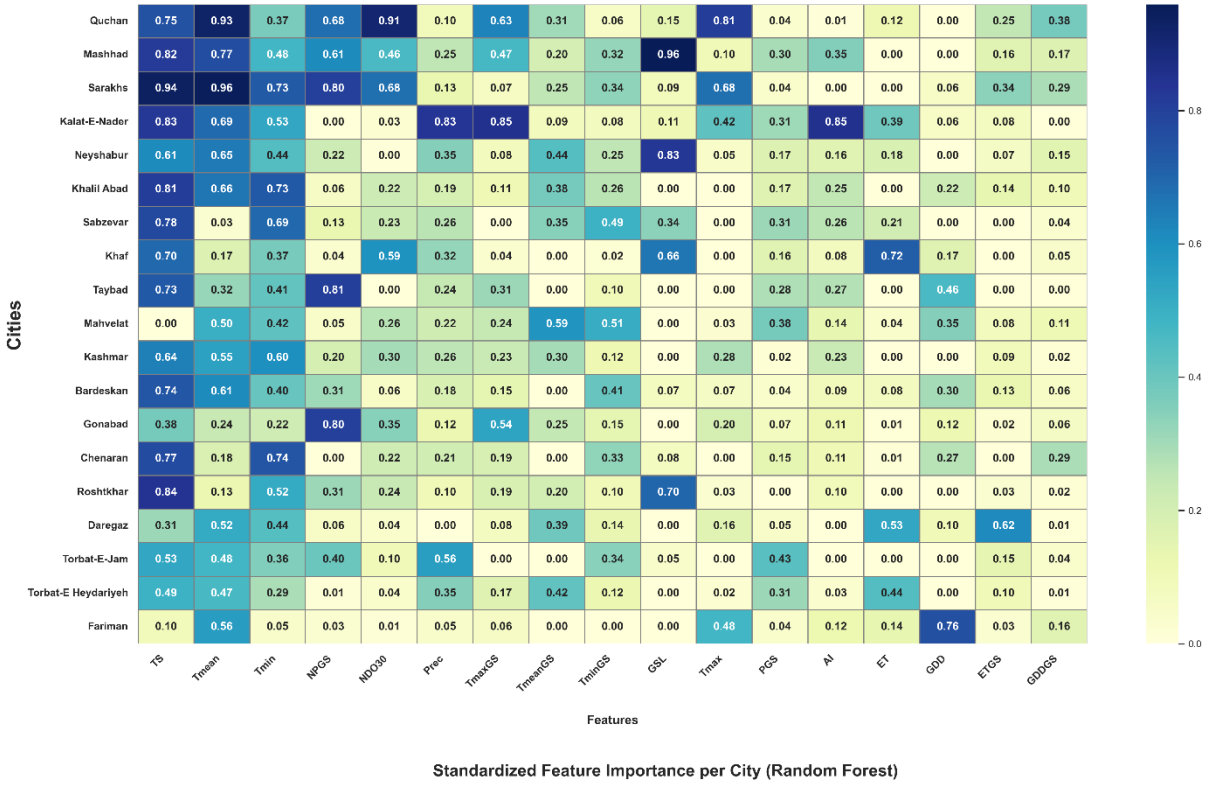
همچنین، دمای کمینه روزانه در هر دو مدل، همبستگی بالایی با سایر متغیرهای دمایی نشان داد؛ موضوعی که بر نقش کلیدی این متغیر در عملکرد مدل‌ها دلالت دارد. با توجه به این یافته‌ها، توجه به دمای کمینه روزانه در تصمیم‌گیری‌های مرتبط با مدیریت زراعی گندم آبی در استان خراسان اهمیت ویژه‌ای دارد.

اهمیت نسبی ویژگی های مورد استفاده در هر الگوریتم در پیش بینی عملکرد گندم

به منظور تحلیل عمیق تر و مختص هر شهرستان، ویژگی های تأثیر گذار در پیش بینی عملکرد گندم به ترتیب اولویت آنها در هر شهرستان، برای هر دو مدل در شکل های ۴ و۵ آورده شده است. داده های این شکل ها مقایسه بین شهرستان ها و شناسایی ویژگی

Iranian Journal of Field Crops Research
Ferdowsi University of Mashhad

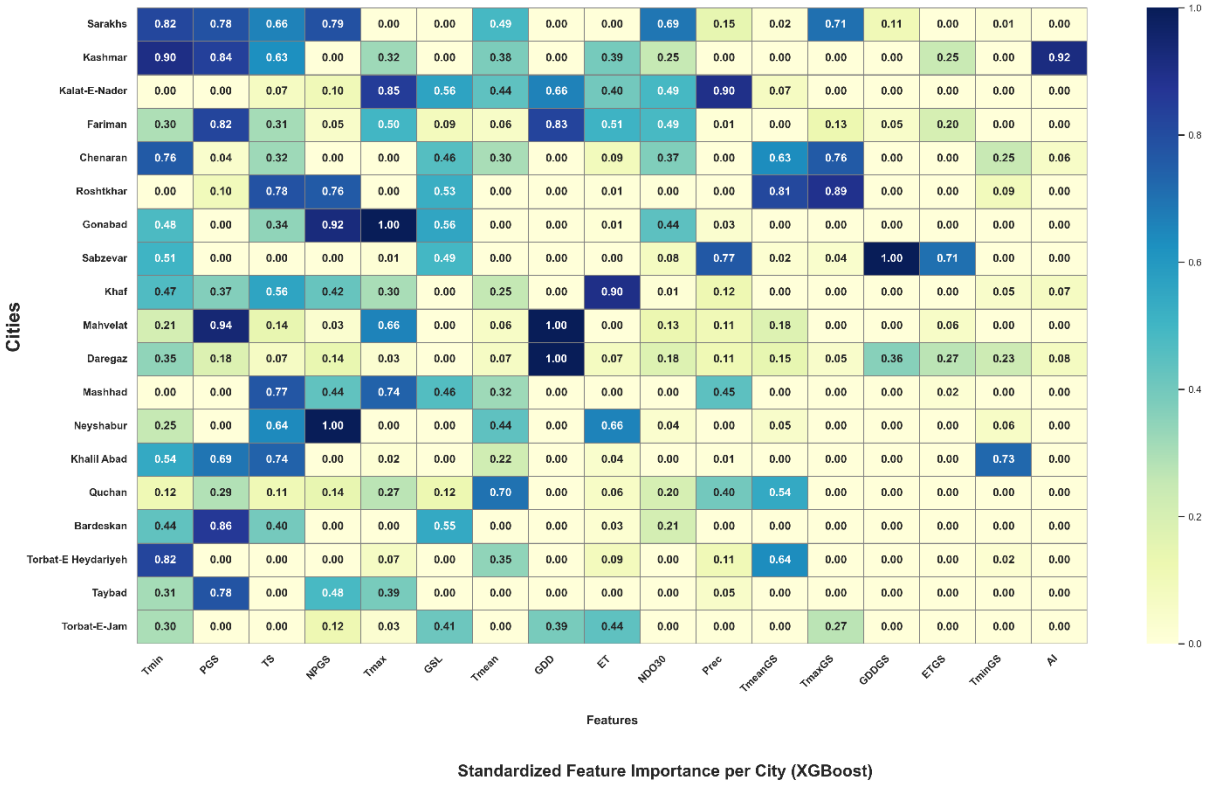
های مشترک و همچنین ویژگی های ایجاد کننده تفاوت را ممکن می سازد. این ویژگی‌ها در مدل‌های مختلف تأثیرگذار هستند و به میزان زیادی پیش‌بینی‌ها را تحت تأثیر قرار می دهند. در رندوم فارست، پنج ویژگی مهم به ترتیب عبارت بودند از: دمای میانگین روزانه، تعداد روزهای با دمای بالای ۳۰ درجه سانتی گراد، دمای بیشینه روزانه، شدت فصلی بودن درجه حرارت، و تعداد دفعات وقوع بارندگی طی فصل رشد. پنج ویژگی برتر در مدل ایکس‌جی‌بوست به شرح زیر بودند: دمای کمینه روزانه، تعداد دفعات وقوع بارندگی طی فصل رشد، بارندگی طی فصل رشد، و دمای بیشینه طی فصل رشد.



شکل ٤– اهمیت نسبی ویژگی های مؤثر بر پیش بینی عملکرد گندم در هر شهرستان توسط مدل رندوم فارست (بر مبنای داده های استاندارد شده). (شرح کامل ویژگی ها در جدول ۲ بخش مواد و روش ها آورده شده است).

Figure 4 – Relative importance of features influencing wheat yield prediction in each county using the Random Forest model (based on standardized data). (A complete description of the features is provided in Table 2 in the Materials and Methods section).

Iranian Journal of Field Crops Research
Ferdowsi University of Mashhad



شکل ۵- اهمیت نسبی ویژگی های مؤثر بر پیش بینی عملکرد گندم در هر شهرستان توسط مدل ایکس‌جی‌بوست (بر مبنای داده های استاندارد شده). (شرح کامل ویژگی ها در جدول ۲ بخش مواد و روش ها آورده شده است).

Figure 5 – Relative importance of features influencing wheat yield prediction in each county using the XGBoost model (based on standardized data). (A complete description of the features is provided in Table 2 in the Materials and Methods section).

در پهنه بندی احتمال رخداد بیماری فوزاریوم گندم با استفاده از روش جنگل تصادفی و داد های حاصل از اینترنت اشیاء، پنج شاخص با درجه اهمیت نسبی زیاد تشخیص داده شدند (مثل تعداد ساعت هایی که دما بین ۱۵ تا ۳۰ درجه سانتی گراد و رطوبت نسبی بیشتر از ۸۰ درصد است). نتایج ارزیابی این مدل حاکی از کارآیی ان در پیش بینی بیماری فوزاریوم گندم بود (Khodabendehloo et al., 1400).

تحلیل بصری ویژگی‌های مؤثر مدل‌ها در سطح شهرستان ها

برای کلیه شهرستان ها نمودارهای ساختار درختی، شیپ، فایل های پی کی ال و معیارهای ارزیابی دقت مدل پیش بینی کننده محاسبه و تولید شدند. به دلیل تعداد زیاد این نمودار ها، در اینجا تنها به ارائهٔ نمودارهای مربوط به شهرستان مشهد به عنوان نمونه اکتفا می شود.

مصورسازی ساختار درختی[1] مدل ایکس‌جی‌بوست و رندوم‌فارست برای شهرستان مشهد

در مدل‌های رندوم‌فارست به دلیل اینکه چندین درخت تصمیم[2] به‌طور همزمان برای پیش‌بینی ساخته می‌شوند، و نیز هر درخت تصمیمات را به‌طور مستقل بر اساس ویژگی‌های داده‌ها می‌گیرد، بنابراین، نتیجه‌گیری‌های مدل، تجمیعی از همه درخت‌هاست، نه یک درخت واحد. به عبارت دیگر، درخت‌های مدل به‌طور تصادفی ساخته می‌شوند و با ایجاد تنوع، قدرت پیش‌بینی مدل بیشتر

[1] Tree Visualization
[2] decision trees

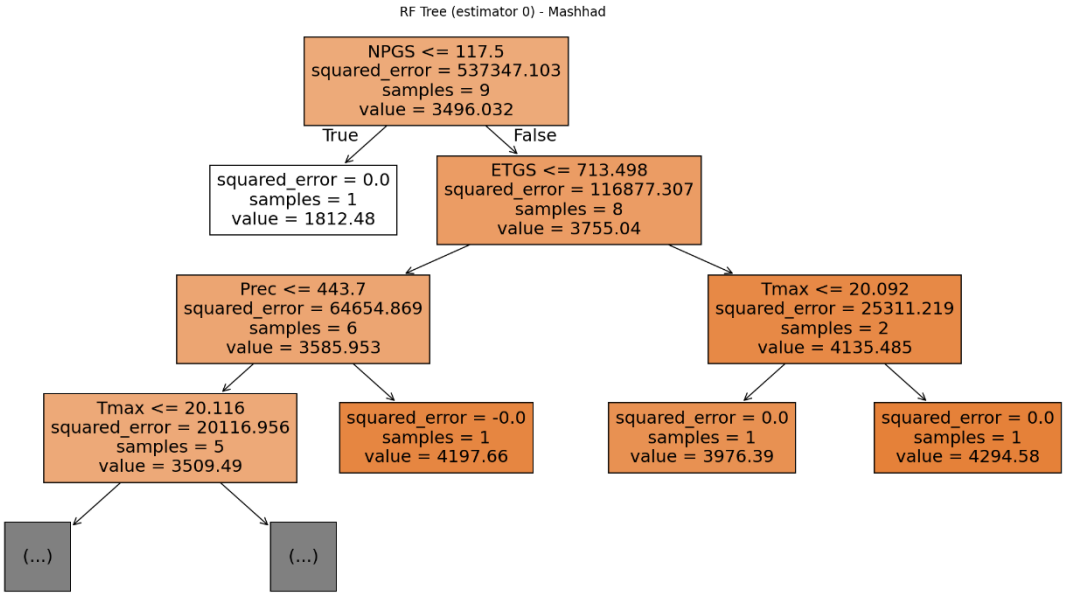
می‌شـــود، پس نمی‌توان یک درخت واحد را به‌عنوان "درخت نهایی" انتخاب کرد، چون همه درخت‌ها در کنار هم نتیجه کلی را تعیین می‌کنند. بنابراین، در مواقعی که نیاز باشد تصویری از ساختار تصمیم‌گیری درخت‌ها ارائه گردد، معمولاً تصویر درخت‌های فردی (برای یک درخت واحد) نمایش داده می شـــود، زیرا نمایش همه درخت‌ها به‌طور مســـتقیم از طریق یک نمودار دشـــوار و بی‌معنا می‌شـود. (اگر تعداد درخت‌ها زیاد باشـد و در حال بررسـی آخرین درخت، یک سـاختار درختی ناقص یا خیلی سـاده در خروجی دیده شود، دلیل آن ممکن است این باشد که آن درخت به دلیل نمونه گیری تصادفی، اطلاعات مفیدی تولید نکرده باشد. درخت‌های رندوم فارســـت برخلاف درخت‌های تصـــمیم منفرد، ممکن اســـت به دلیل نمونه‌گیری تصـــادفی ویژگی‌ها[1] در برخی درخت‌ها مسیرهای خیلی کوتاه ایجاد کنند، مثلاً حالتی که در یک درخت، فقط یک یا دو ویژگی مؤثر بوده‌اند و بقیه مسیر تقسیم ســـاخته نشـــده باشـــد). حالت دیگر، برگ نهایی زودرس[2] نامیده می شـــود به این معنی که در بعضـــی درخت‌ها (به‌ویژه در رندوم فارســـت)، ممکن اســـت برخی گره‌ها زودتر به برگ تبدیل شـوند، زیرا همه نمونه‌ها در آن گره از یک کلاس هسـتند، یا به حداقل تعداد نمونه برای تقسیم[3] یا حداقل نمونه برای برگ[4] رسیده‌اند. در این صورت، آن گره دیگر تقسیم نمی‌شود، و به صورت برگ (حتی در عمق کم) باقی می‌ماند. این موضـوع طبیعی و بخشـی از فرآیند یادگیری درخت اسـت. راه حل این مشـکل، اسـتفاده از نمودارهای اهمیت ویژگی[5] است که می‌توانند به‌صورت گرافیکی تاثیر هر ویژگی بر پیش‌بینی مدل را نشان دهند (شکل ۷). با این حال، برخی محققان پیشنهاد کرده اند که نمایش آخرین درخت نیز به دلیل ارائه اطلاعات و ویژگی های نهایی شناسایی شده، می تواند مفید باشد (شکل ۶آ).

در مدل ایکس‌جی‌بوســـت، چون مدل به‌طور تدریجی و گام‌به‌گام درخت‌ها را اضـــافه می‌کند، اولین درخت، بیشـــترین تاثیر را بر پیش‌بینی‌های مدل دارد، زیرا بقیه درخت‌ها در مسیر اصلاح خطاهای قبلی عمل می‌کنند. مفهوم بوستینگ[6] در ایکس‌جی‌بوست و دیگر مدل‌های مبتنی بر بوســـتینگ، به همین مفهوم اشـــاره دارد، یعنی مدل به تدریج (درخت به درخت) یاد می‌گیرد و درخت‌های بعدی بیشتر به اصلاح پیش‌بینی‌های غلط (خطا) در درخت‌های قبلی می‌پردازند. درخت‌های ابتدایی (از نظر رتبه) اطلاعات بیشتری ارائه می‌دهند. درخت اول اطلاعات اولیه‌ای از روابط پیچیده داده‌ها به مدل می‌دهد و می‌توان آن را به‌صورت بصری بررسی کرد. بنابراین، مزایای نمایش اولین درخت در ایکس‌جی‌بوست به شرح زیر هستند: نمایش ساده‌تر و قابل درک‌تر از تصمیمات مدل، خصـــوصـــاً درخت اول که بیشـــترین تغییر را ایجاد می‌کند؛ می‌تواند به‌عنوان یک نقطه شـــروع خوب برای توضـــیح مدل و نحوه تصـــمیم‌گیری‌های اولیه باشـــد؛ به‌ویژه برای درک ویژگی‌های کلیدی که در درخت اول مورد توجه قرار گرفته‌اند، می‌تواند مفید باشد.

در نمودارهای درختی نهایی، ساختار تصمیم‌گیری مدل‌ها بر اساس شاخص‌های اصلی، از جمله ویژگی‌هایی نظیر میزان بارندگی، دمای هوا، تاریخ کاشت، و غیره نمایش داده شده است. مدل ایکس‌جی‌بوست و رندوم‌فارست هر دو درخت‌هایی با ساختار نسبتاً ساده برای شهرستان مشهد تولید کردند، ایکس‌جی‌بوست با ۴ سطح و رندوم‌فارست با ۴ سطح، (شکل ۶ آ و ۶ ب)، اما تفاوت‌هایی اساسی در روش شناختی، ساختار و نوع متغیرهای به‌کاررفته و نحوه تقسیم نمونه‌ها دارند که عبارتند از:

[1] feature bagging
[2] Early Leaf Node
[3] min_samples_split
[4] min_samples_leaf
[5] Importances of Features
[6] Boosting

متغیرهای تأثیرگذار اولیه: به‌منظور درک ساختار تصمیم‌گیری مدل رندوم‌فارست، درخت نهایی از میان درخت‌های ساخته‌شده در مدل بهینه شده برای هر شهرستان ترسیم شد. هرچند تصمیم نهایی مدل به‌صورت تجمیعی از تمام درخت‌ها حاصل می‌شود، اما نمایش درخت نهایی می‌تواند دید مناسبی از مسیرهای تصمیم‌گیری مدل نسبت به متغیرهای ورودی فراهم کند (مانند آستانه‌های بحرانی و اثرات متقابل ویژگی‌ها). در مدل رندوم‌فارست، متغیر تعدا دفعات وقوع بارندگی طی فصل رشد، نقش اساسی در پیش بین عملکرد گندم دارد. در مدل رندوم فارست ، متغیر دوم باز هم ماهیت رطوبتی دارد (تبخیر و تعرق طی فصل رشد).



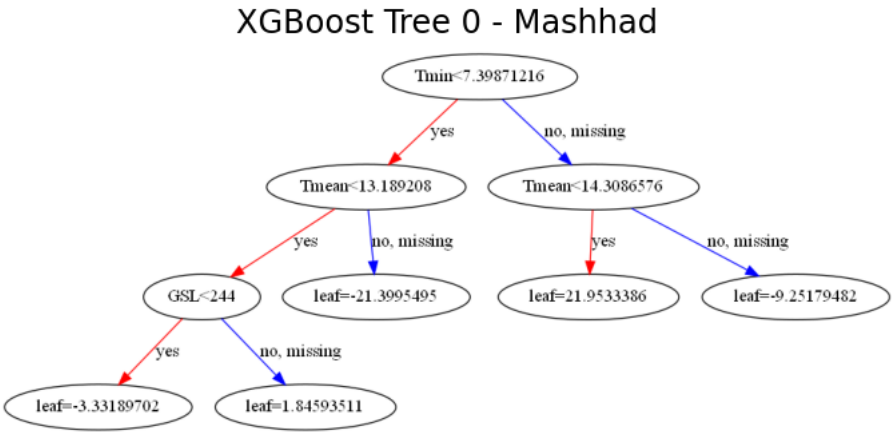
شکل ٦a : آخرین نمودار درختی تولید شده توسط مدل رندوم‌فارست برای شهرستان مشهد. تعداد دفعات وقوع بارندگی طی فصل رشد NPGS، تبخیر و تعرق طی فصل رشد ETGS، بارندگی PREC، دمای بیشینه روزانه Tmax .

Figure 6A – The final decision tree generated by the Random Forest model for Mashhad county. Number of rainy days during the growing season (NPGS), evapotranspiration during the growing season (ETGS), precipitation (PREC), and daily maximum temperature (Tmax)

در مدل ایکس‌جی‌بوست، متغیر دمای کمینه روزانه اولین متغیر تفکیک‌کننده بوده که دارای بالاترین میزان گین[1] یا برتری است و نقش اساسی در کاهش خطای مدل ایفا می‌کند. این موضوع نشان می‌دهد که در چارچوب این مدل، دما نقش محوری در پیش‌بینی عملکرد گندم داشته است. ویژگی دوم در مدل ایکس‌جی‌بوست ، دوباره همان متغیر دمایی است (دمای میانگین روزانه).

[1] Gain

Iranian Journal of Field Crops Research



شکل ٦b : اولین ساختار درختی تولید شده توسط مدل ایکس‌جی‌بوست برای شهرستان مشهد. دمای کمینه روزانه Tmin، دمای میانگین روزانه Tmean، طول فصل رشد GSL.

Figure 6B – The first tree structure generated by the XGBoost model for Mashhad county. Daily minimum temperature (Tmin), average daily temperature (Tmean), and growing season length (GSL)

رفتار مدل در تفکیک داده‌ها: در مدل رندوم‌فارست، از ۹ نمونه، ۸ نمونه (یعنی ۸۸/۸ درصد) به شاخه دوم منتقل می‌شوند و سپس ۲و ۶ نمونه به برگ های بعدی و در نهایت یک نمونه در برگ نهایی باقی می‌مانند، یعنی مدل در تقسیم‌های عمیق‌تر فقط روی نمونه‌های کمتری تمرکز کرده است. این موضوع گاهی به بیش‌برازش[1] رندوم‌فارست منجر می‌شود. در مقابل، مدل ایکس‌جی‌بوست با پوشش یکنواخت‌تر و تقسیم‌بندی‌های هدفمندتر، متغیرهایی مانند دمای کمینه روزانه، دمای میانگین روزانه و طول فصل رشد را به‌کار گرفته که هم متنوع‌تر هستند و هم هر یک بر اساس مقدار گین سهم مشخصی در کاهش خطای کلی داشته‌اند.

ساختار مدل، تعداد و مقدار برگ و تفسیر عملکرد: در رندوم‌فارست، مقدار نهایی[2] معادل ۴۲۹۴/۵ واحد عملکرد گندم (کیلوگرم در هکتار) برای یک نمونهٔ خاص است. در ایکس‌جی‌بوست، مقدار نهایی در برگ‌ها معمولاً مقدار اصلاح شده یا باقی‌مانده[3] است، نه مقدار نهایی عملکرد. این موضوع باعث می‌شود که تفسیر مستقیم خروجی درخت سخت‌تر باشد، ولی مدل دقت بیشتری در مجموع خواهد داشت، چراکه این مقادیر به صورت تجمعی در بوسترها لحاظ می‌شوند. در مدل رندوم‌فارست ساختار ساده‌تر و تفسیر مستقیم‌تری ارائه می‌شود، ولی با خطر بیش‌برازش در برگ‌هایی با تعداد کم مواجه است. مدل ایکس‌جی‌بوست گرچه پیچیده‌تر به‌نظر می‌رسد، و معمولاً عمق بیشتری در شاخه‌های مرتبط با ویژگی‌ها ایجاد می کند، اما تقسیم‌بندی‌های آن بر اساس کاهش خطای مؤثرتر، بهره‌گیری از متغیرهای متنوع، و مشارکت تدریجی در تقویت مدل نهایی، قدرت بیشتری در مدل‌سازی دارد. در نهایت، بسته به هدف مطالعه (تفسیر یا دقت پیش‌بینی)، هر یک از دو مدل مزایای خاص خود را دارند، ولی برای مدل‌سازی دقیق‌تر و شناسایی متغیرهای تأثیرگذار کلیدی، تفسیرهای ایکس‌جی‌بوست می توانند اطلاعات عمیق‌تری فراهم می کنند. گزارش شده است که رندوم‌فارست عملکرد نیشکر را بهتر از مدل‌های خطی سنتی برآورد می کند (Everingham et al., 2016).

[1] Overfitting
[2] Leaf value
[3] residual

به طور کلی، ساختارهای درختی بر قابلیت تحلیل و تفسیر نتایج مدل ها بسیار تأثیر گذار هستند و می توانند در مطالعات آینده برای شفاف‌سازی تصمیم‌های مدل به کار روند.

تحلیل نمودارهای اهمیت نسبی ویژگی ها (نمودارهای شیپ ویولونی) مدل رندوم فارست برای شهرستان مشهد

در مدل رندوم‌فارست برای شهرستان مشهد، ویژگی‌هایی که بیشترین تاثیر را در پیش‌بینی عملکرد گندم دارند به ترتیب از بیشترین اهمیت به کمترین اهمیت عبارتند از:

۱. Tminدمای حداقل شبانه نقش مهمی در شدت تنفس نگهداری، تسهیم و توزیع فرآورده های فتوسنتزی تولید شده طی مرحله روشنایی، تعادل هورمونی و غیره دارد. در دوره‌های شبانه، دمای پایین‌تر می‌تواند شدت تنفس را کاهش داده و سبب ذخیره بیشتر انرژی در گیاه شود (Hatfield & Prueger, 2015).

۲. TS این ویژگی دومین رتبه اهمیت را در بین ویژگی ها داشت و تأثیر بسزایی در پیش‌بینی‌ها دادارد. همچنین، نقش برجسته ای در تغییرات دمایی و استرس دمایی (در طول فصل رشد گندم) دارد.

۳. TmeanGSمیانگین دمایی مطلوب سبب تعادل بین فتوسنتز و تنفس شده و به رشد بهینه کمک می‌کند (Asseng et al., 2015).

٤. TminGSدماهای پایین‌تر می‌توانند تأثیر منفی بر جوانه‌زنی و مراحل اولیه رشد داشته باشند، ولی در مواردی نیز از استرس گرمایی جلوگیری می‌کنند.

٥. PGS مقدار بارندگی در طول فصل رشد رتبه پنجم اهمیت در بین ویژگی های مورد بررسی را به خود اختصاص داد. تأمین آب کافی و منظم در دوره رشد برای تأمین آب و جذب عناصر غذایی حیاتی است (Si et al., 2023) .

٦. NPGS توزیع زمانی بارش حتی مهم‌تر از مقدار آن است، زیرا دفعات بارندگی در تثبیت ظرفیت ذخیره رطوبت در خاک نقش کلیدی دارد.

۷. GDD این شاخص، جمع دماهای مفید برای رشد است و با مراحل فنولوژیک گیاه ارتباط مستقیم دارد (Vieira et al., 2025).

۸. Prec بارش‌ها بر آب‌رسانی به گیاه و تأثیر مستقیم بر برآیندهای رشد و نمو و در نهایت تعیین میزان عملکرد نهایی تأثیر مستقیم دارند. تأمین نیاز آبی گیاه، بر مراحل حساس رشد از جمله خوشه‌دهی و پر شدن دانه تأثیر می‌گذارد. یافته‌ها تأکید می‌کنند که حتی در شرایط آبیاری، مدیریت مناسب زمان‌بندی و میزان آبیاری (بارندگی) برای بهینه‌سازی جذب عناصر غذایی و افزایش عملکرد گندم ضروری است (Wang and Li, 2002).

۹. Tmeanدمای میانگین روزانه در طول دوره رشد تأثیر زیادی بر عملکرد گندم دارد. این ویژگی به ویژه در نواحی با تغییرات دمایی زیاد مهم است.

۱۰. NDO30 تعداد روزهایی که دما بیش از ۳۰ درجه سانتی‌گراد است، نشان‌دهنده شرایط دمایی بحرانی برای گیاه می‌باشد (Farhadi et al., 2024).

۱۱. Tmax افزایش بیش از حد این دما، فرآیند فتوسنتز را مختل و تعرق را تشدید می‌کند که منجر به افت عملکرد می‌شود.

۱۲. ETGS یکی از شاخص‌های کلیدی روابط آبی بوده که ارتباط نزدیکی با نیاز آبی گیاه دارد.

۱۳. AI شاخص خشکی، بر عملکرد تأثیر منفی داشته است که این موضوع بسیار بدیهی است و توسط متخصصان کشاورزی برای ارزیابی شرایط عمومی یک ناحیه برای فعالیت های کشاورزی استفاده می‌شود.

۱٤. TmaxGS باعث بروز تنش گرمایی در طول فصل رشد شده و رشد سلولی و پر شدن دانه را مختل می‌کند.

۱٥. ET در مناطق گرم و خشک، این شاخص معیار مناسبی برای بررسی شدت نیاز آبی و خطر تنش خشکی است.

۱۶. GDDGS مشابهGDD ، این ویژگی در تحلیل مراحل نموی از نظر زمان وقوع و طول دوره هر مرحله نموی عامل تعیین کننده ای به شمار می رود به‌طور خاص نشان‌دهنده تأثیر گرمای تجمعی در فصل رشد بر پیشرفت فنولوژیک است.

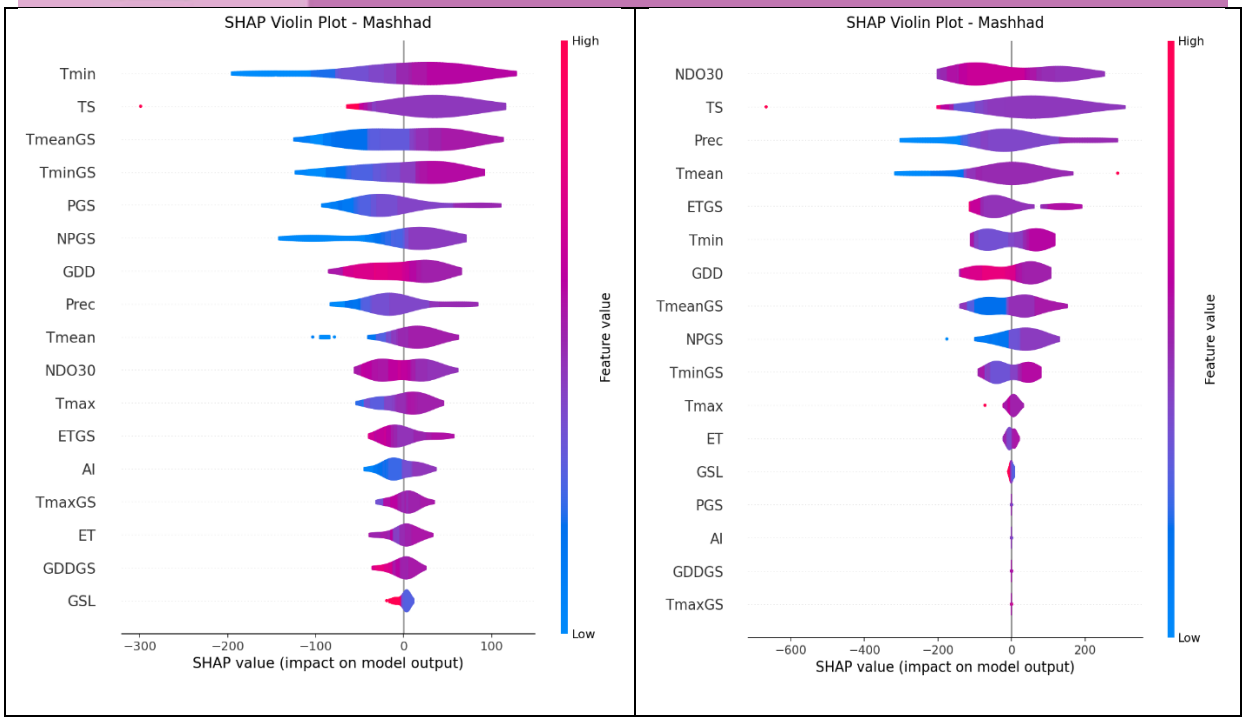
۱۷. GSL رابطه مستقیمی با فرصت زمانی برای فتوسنتز و پر شدن دانه دارد؛ طولانی‌تر بودن این فصل معمولاً عملکرد بالاتر را سبب می‌شود.

ویژگی های با رتبه های ۱ تا ۵، به ویژه در مناطق با تغییرات دمایی زیاد، شـــرایط رشـــد گیاه را تحت تأثیر قرار می‌دهند و مدل به خوبی اهمیت آنها را پیش بینی کرده است (Tize et al., 2018; Slafer et al., 2023). نتایجی مشابه با نتایج مطالعه حاضر توسط Soleimannejad et al., (2018) گزارش شده است مبنی بر این که الگوریتم ناپارامتریک رندوم فارست روش مناسبی در برآورد مشخصه های سه مقطع، تاج پوشش و تراکم جنگل های زاگرس با استفاده از تصاویر OLI ماهواره لندست بود. در پژوهشی، با دو رهیافت ارزیابی میدانی و سنجش از دور، بین نتایج حاصل از بکارگیری دو مدل رگرسیون خطی چندگانه و جنگل تصـــادفی در پیش بینی نیتروژن کل خاک، دو مدل اختلاف معنی داری با رهیافت ارزیابی میدانی نداشـــتند (Sadeghi & Ahmadi Nadoushan, 1400).

تحلیل نمودارهای اهمیت نسبی ویژگی ها (نمودارهای شیپ ویولونی) مدل ایکس‌جی‌بوست برای شهرستان مشهد

داده های حاصل از مدل ایکس‌جی‌بوست برای شهرستان مشهد، ترتیب اهمیت متغیرها را متفاوت با مدل رندوم فارست نشان داد، با بیان این نکته مهم که در عین حال یک متغیر مهم این مدل در گروه پنج متغیری مهم مدل رندوم فارســـت نیز حضـــور دارند (به عبارت دیگر، اهمیت نسبی شدت فصلی بودن درجه حرارت را به خوبی تشخیص داده است) (شکل ۷). در این مدل اولین و مهم ترین ویژگی، تعداد روزهای با دمای بالاتر از ۳۰ درجه ســانتی گراد تشــخیص داده شــد. در رتبه دوم شــدت فصــلی بودن درجه حرارت، بارندگی در رتبه سوم و در رتبه چهارم مشابه با مدل رندوم فارست، متغیر دمایی میانگین روزانه (در رندوم فارست دمای میانگین طی فصل رشد است) قرار گرفت. این ویژگی‌ها ارتباط مستقیمی با رشد گیاهان بویژه در مناطق گرم و خشک دارند و به همین دلیل در مدل ایکس‌جی‌بوست، به خصوص در مناطقی که گندم به میزان بیشتری تحت تأثیر گرما و خشکی که مستقیماً رشد و فیزیولوژیی گیاه را تحت تأثیر قرار می دهند، اهمیت بیشتری پیدا کرده اند.

Iranian Journal of
Field Crops Research



شـــکل ۷- نمودارهای شـــیپ (ویولونی) اهمیت نســـبی ویژگی ها در مدل های بهینه شـــده رندوم فارســـت (ســـمت چپ) و ایکس‌جی‌بوست (سمت راست) در پیش بینی عملکرد گندم در شهرستان مشهد. (شرح کامل ویژگی ها در جدول ۲ بخش مواد و روش ها آورده شده است).

Figure 7 – SHAP (violin) plots of the relative importance of features in the optimized Random Forest model (left) and XGBoost model (right) for predicting wheat yield in Mashhad county. (A complete description of the features is provided in Table 2 in the Materials and Methods section).

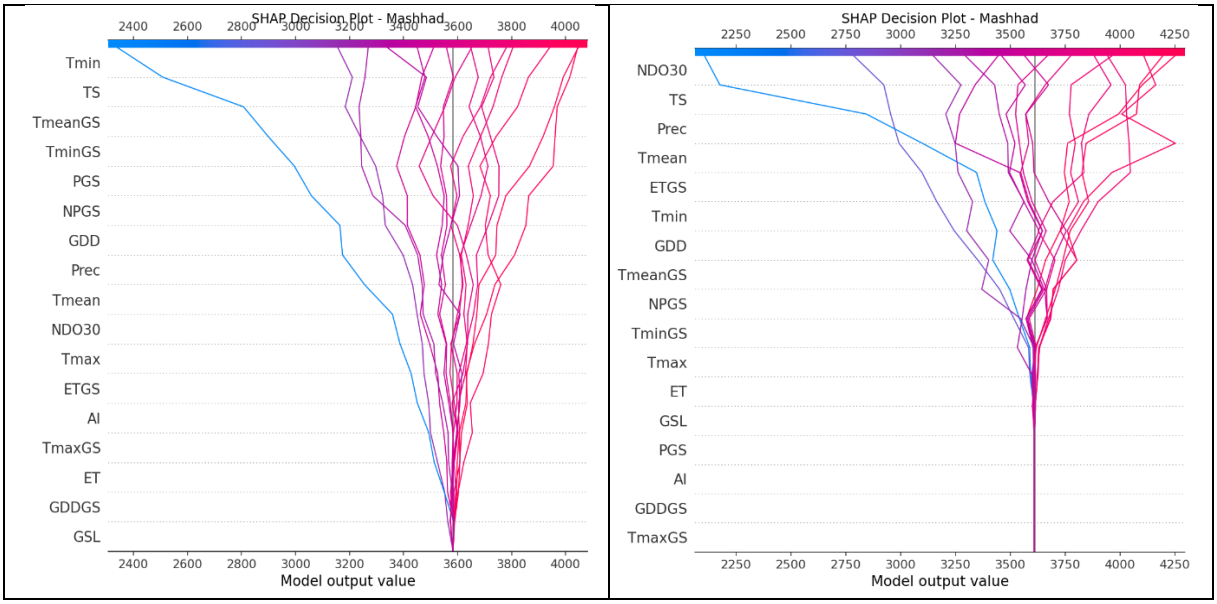
نکته قابل توجه اینکه در هر دو مدل، متغیرهای با ماهیت دمایی حائز بیشـترین اهمیت نسـبی تشـخیص داده شـده اند. ویژگی های بعدی (از رتبه دهم تا هفدهم) نسبت به مدل رندم فارسـت، اهمیت بسـیارکمتری را به خود اختصـاص داده اند. در تحلیل و تفسـیر نتایج اینچنینی، باید دقت شـود تا از دلیل تراشـی بی مورد و توجیه های غیر منطقی پرهیز گردد. برای مثال، اهمیت عامل بارندگی، در مدل ایکس جی بوسـت در رتبه سـوم تعیین شـده اسـت. همین متغیر، در مدل رندوم فارسـت در رتبه هشـتم قرار گرفته اسـت (پراکنش مقادیر شیپ آن در هر دو مدل تقریبا یکسان است). تأکید بیش از حد بر اهمیت بارندگی در مدل ایکس جی بوست، می تواند نقش تبین کننده عوامل دمایی مشترک در هر دو مدل را تحت الشعاع قرار دهد. به عبارت دیگر، عدم قطعیت موجود در مدل ها و تفاوت شـرایط واقعی با آنچه که در تئوری فرض می شـود، در نظر گرفته شـود و به یاد داشـت که معنی داری آماری، و معنی داری بیولوژیکی دو مفهوم کاملاً متفاوت هستند.

بارندگی در هر دو مدل اهمیت دارد، اما در جایی که این ویژگی به عنوان یک فاکتور کلیدی در پیش‌بینی‌ها دیده شـود، تأثیرات آن در مدل ایکس‌جی‌بوست ممکن است بیشتر از مدل رندوم‌فارست باشد.

تحلیل نمودارهای شیپ تصمیم و وابستگی جزئی مدل رندوم فارست برای شهرستان مشهد

شکل ۸ یک نمودار تصمیم مبتنی بر مقادیر شیپ برای مدل رندوم فارست (سمت چپ) و ایکس جی بوست (سمت راست) است که برای پیش‌بینی عملکرد گندم در شهرستان مشهد مورد استفاده قرار گرفته است. مقادیر شیپ ابزار قدرتمندی برای درک نحوه‌ی اثرگذاری هر ویژگی ورودی بر پیش‌بینی مدل هستند. در این تصویر، تغییرات مقادیر شیپ برای هر ویژگی نسبت به مقادیر آن

ویژگی در نمونه‌های مختلف مشاهده می‌شود. در این نمودار، محور افقی نمایانگر مقدار شیپ است، که میزان تأثیر هر ویژگی را بر پیش‌بینی عملکرد نشان می‌دهد (مثلاً افزایش یا کاهش عملکرد). محور عمودی ویژگی‌های مدل را فهرست کرده است (مانند دمای کمینه روزانه، شدت فصلی بودن درجه حرارت، دمای میانگین طی فصل رشد و غیره). هر خط بیانگر یک نمونه (سال-مکان[1] در این جا منظور عملکرد گندم در یک سال خاص در مشهد است) است. هر کدام از این خطهای رنگی هنگام عبور از یک ویژگی (مثلا دمای کمینه روزانه)، به سمت چپ یا راست منحرف می‌شوند: اگر به سمت چپ برود، یعنی شیپ منفی بوده و عملکرد کاهش می یابد. اگر خط به سمت راست خط پایه منحرف شود، یعنی شیپ مثبت بوده و عملکرد افزایش یافته است. در ویژگی‌هایی که قدرت توضیح‌دهندگی بالایی دارند، این انحراف‌ها زیاد و شدید هستند. اگر خطوط در یک ویژگی به‌صورت قابل توجهی از هم جدا بشوند (باز شوند)، یعنی آن ویژگی اثر زیادی در خروجی مدل دارد. اگر خطوط تقریباً صاف یا هم‌راستا باقی بمانند، آن ویژگی اثر چندانی در مدل نداشته است. اثر یک ویژگی همیشه خطی یا یکنواخت نیست؛ یعنی مقدار بیشتر یک ویژگی لزوماً همیشه به معنای بهتر یا بدتر بودن نیست، بلکه ممکن است از یک نقطه به بعد، اثر آن ویژگی کاهش پیدا یافته یا حتی برعکس شود. این پدیده در مدل‌های پیچیده مثل رندوم فارست یا ایکس جی بوست به‌وضوح دیده می‌شه و می‌تواند در تحلیل شیپ یا نمودارهای تصمیم مشخص بشود.



شکل ۸- نمودارهای شیپ تصمیم برای پیش بین عملکرد گندم در شهرستان مشهد توسط مدل رندوم فارست (سمت چپ) و ایکس جی بوست (سمت راست)

Figure 8 – Decision SHAP plots for predicting wheat yield in Mashhad county by the Random Forest model (left) and XGBoost model (right)

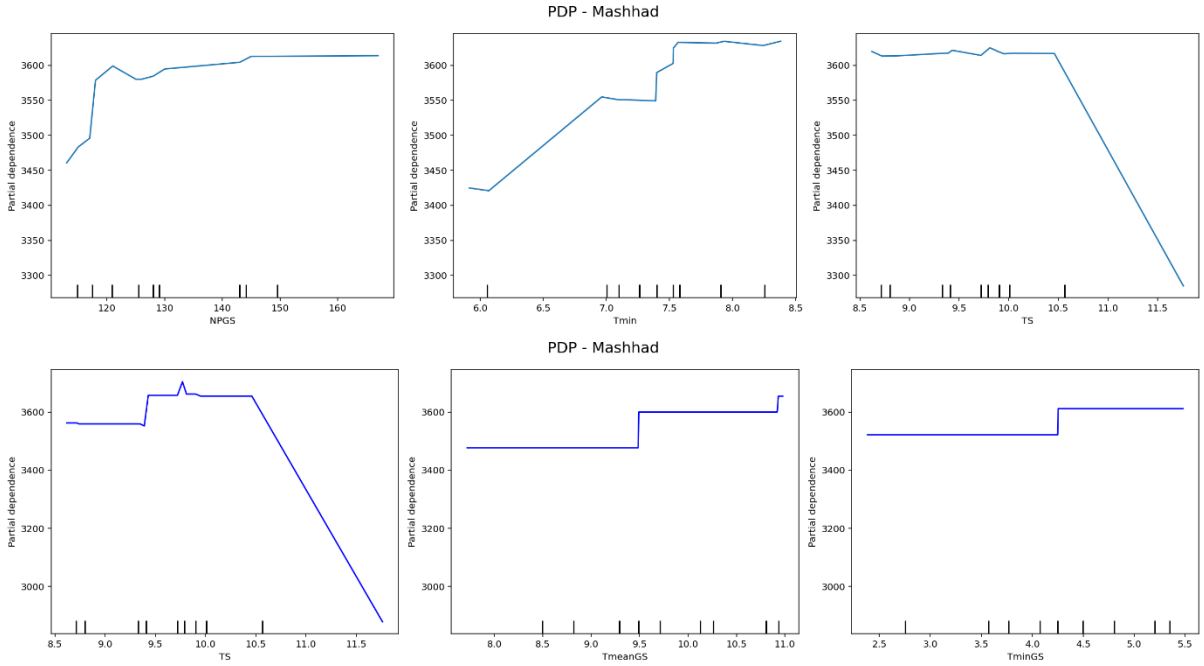
همانطور که در شکل ۸ (سمت چپ) مشاهده می شود، دمای کمینه روزانه در رأس ویژگی‌های مؤثر قرار دارد، به عبارت دیگر، اولین ویژگی واردشونده در مدل است؛ یعنی مدل رندوم فارست در ابتدا تصمیم خود را براساس مقدار دمای کمینه روزانه در هر نمونه شکل می‌دهد. در شکل ۸ خط هایی که از ویژگی دمای کمینه روزانه عبور می کنند، به‌شدت به چپ یا راست منحرف شده اند (قرمزها به راست (عملکرد بالا) یعنی دمای کمینه روزانه بالا، باعث افزایش عملکرد شده؛ آبی‌ها به چپ (عملکرد پایین) یعنی دمای کمینه روزانه پایین باعث افت عملکرد شده اند. بسیاری از نمونه‌ها، مقادیر پایین دمای کمینه روزانه (آبی‌رنگ) باعث کاهش

[1] year-location

عملکرد شده‌اند (سمت چپ خط متوسط عملکرد یا مقدار پایه ۳۶۰۰ کیلوگرم در هکتار قرار دارند؛ به عبارت دیگر، شیپ این مقدار پایین منفی بوده، یعنی باعث کاهش عملکرد نسبت به میانگین مدل شده)،که بیانگر حساسیت گیاه به دماهای پایین در مراحل ابتدایی رشد است.

برای سایر ویژگی ها، به طور مشابه می توان تحلیل ارائه داد. بطور کلی، مدل رندوم فارست توانسته رابطه پیچیده‌ای بین متغیرهای اقلیمی و عملکرد گندم استخراج کند. نمودار شیپ تصمیم نشان می‌دهد که دمای کمینه روزانه، شدت فصلی بودن درجه حرارت، دمای مینگین طی فصل رشد، دمای میانگین روزانه، و بارندگی طی فصل رشد مهم‌ترین نقش را در در پیش‌بینی عملکرد برای مشهد دارند. این تحلیل برای کشاورزان و سیاست‌گذاران می‌تواند در بهینه‌سازی تاریخ کاشت و آبیاری مؤثر باشد. نمودار تصمیم در الگوریتم ایکس جی بوست مشهد در سمت راست شکل ۸ نشان داده شده است.

شکل ۹ که از کاربردی ترین خروجی های شیپ محسوب می شود، نمودار وابستگی جزئی یا دی پی دی نام دارد که اثر مستقیم و جداگانه هر ویژگی بر عملکرد گندم را نمایش می‌دهد، در حالی که سایر ویژگی‌ها ثابت نگه داشته شده‌اند. در ساختار این نمودار، روی محور افقی مقدار ویژگی(مثلاً دمای میانگین روزانه، دمای میانگین طی فصل رشد یا غیره)، روی محور عمودی، مقدار پیش‌بینی‌شده عملکرد (توسط مدل) قرار می گیرند. خطوط نمودار، منحنی پاسخ مدل نسبت به تغییرات هر ویژگی را نشان می دهد. در مدل رندوم فارست (شکل بالایی)، افزایش تعداد دفعات وقوع بارندگی از ۱۱۲ تا ۱۳۰ مرتبه، مقدار عملکرد گندم در مشهد با شیب زیادی افزایش می یابد و به حدود ۳۶۰۰ کیلوگرم در هکتار می رسد، ولی پس از آن با افزایش دفعات بارندگی تا حدود ۱۷۰، این شیب افزایشی کاهش یافت و به حالت تقریباً اشباع درآمد (یعنی افزایش بیشتر این ویژگی تأثیر محسوسی بر عملکرد ندارد). افزایش دمای کمینه روزانه، از ۶ تا ۸/۵ درجه سانتی گراد، سبب افزیش یکنواخت عملکرد گندم شد. این روند افزایشی نشان می‌دهد که در شرایط آب‌وهوایی مشهد، گندم نسبت به افزایش دمای کمینه تا این بازه پاسخ مثبت دارد. با این توضیح که در محدودهٔ ۷ تا ۷/۵ درجه سانتی گراد، نوسانات جزئی وجود داشت که احتمالاً به اثرات پیچیده تعامل بین ویژگی‌ها یا نویز داده مربوط است.



شکل۹ – نمودار شیپ وابستگی جزئی (دی پی دی) برای مدل رندوم فارست (بالا) و ایکس جی بوست (پایین) در پیش بینی عملکرد گندم مشهد

Figure 9 – SHAP Partial Dependence (DPD) plots for the Random Forest model (top) and XGBoost model (bottom) in predicting wheat yield in Mashhad

افزایش شاخص خشکی در بازه ۸/۵ تا حدود ۱۰/۵، اثر منفی چندانی ندارد، حتی گاهی عملکرد اندکی بیشتر هم می‌شود، اما پس از حدود ۱۰/۵، روند تغییر ناگهانی شده و با افزایش مقدار شاخص شدت فصلی بودن درجه حرارت به بیش از ۱۱/۵، عملکرد گندم حدود ۳۰۰ کیلوگرم در هکتار افت می‌کند که این یک نقطه آستانه بحرانی است که از آن به بعد خشکی به طور محسوس به عملکرد آسیب می‌زند. یافته‌ها تأیید می‌کنند که تعادل در گرما(دمای کمینه روزانه)، شاخص خشکی و شرایط مناسب آبی(تعداد دفعت وقوع بارندگی طی فصل رشد) برای دستیابی به عملکرد بهینه گندم در مشهد ضروری هستند. همچنین وجود خشکی شدید (شاخص خشکی)یک تهدید محسوب می‌شود.

نتیجه‌گیری: در مطالعات متعدد بیان شده است که مدل ایکس‌جی‌بوست به ویژگی‌های خاص‌تر و پیچیده‌تر مانند درجه روز رشد توجه دارد، و می‌تواند تعاملات بیشتری بین ویژگی‌ها را مدل سازی کند، اما برای شهرستان مشهد اینچنین نبود. این نمودارها به کارشناسان زراعت کمک می‌کنند تا دلیل تصمیم‌گیری مدل‌ها را بهتر درک کنند و از آن برای بهینه‌سازی کشت و مدیریت مزرعه در سال‌های آینده بهره ببرند. نمودارهای تکمیلی برای سایر شهرستان‌ها تولید شده و بنا به درخواست در دسترس علاقمندان قرار می گیرد.

گزارش شده است که الگوریتم رندوم فارست در ارزیابی توان اکولوژیک کاربری جنگلداری در حوزه آبخیز کن (استان تهران)، توان بسیار بالایی در پیش بینی مناطق مستعد از خود نشان داد و متغیرهای عمق خاک، بارندگی در فصل رویش، ارتفاع، بافت خاک، شیب و جها به ترتیب دارای اهمیت از یاد به کم بودند (Oghnoum et al., 2019). در بکارگیری مدل رگرسیونی جنگل تصادفی به منظور برآورد زیتوده روی زمین جنگل با استفاده از تصاویر ماهواره ای، مشخص شد که شاخص های RVI، NDVI،GNDVI و DVI اهمیت بیشتری در ارائه مدل برآورد زیتوده داشتند. ضریب تبیین این مدل ۸۰/. و جذرمیانگین مربعات خطا برابر با ۲۸/۷ گزارش شد (Gheysabeigi et al., 2024).

ارزیابی و تعیین اعتبار

به منظور تحلیل عمیق تر و مقایسه دقیق تر دقت دو الگوریتم در پیش بین عملکرد گندم، چهار معیار ارزیابی مدل ها به تفکیک برای ۱۹ شهرستان، و همچنین میانگین معیارهای ارزیابی دو الگوریتم، در جدول ۵ آورده شده اند. ترتیب شهرستان ها بر مبنای معیارهای ارزیابی، در هر دو الگوریتم و نیز میانگین آنها، از زیاد به کم، مرتب شده اند.

- **جذرمیانگین مربعات خطا:** هرچه این مقدار کمتر باشد، مدل پیش‌بینی دقیق‌تری دارد. درصد تفاوت برای RMSE نشان‌دهنده این است که کدام مدل به طور متوسط خطای بزرگتری دارد. برای بیشتر شهرستان‌ها، مدل ایکس‌جی‌بوست خطای بالاتری دارد. در این بین، چند مورد وجود دارد که این مدل عملکرد بهتری نشان داده است (مثل شهرستان‌های خاف، مشهد، رشتخوار و تایباد).

جدول ۵- چهار معیار ارزیابی دو الگوریتم رندوم فارست و ایکس جی بوست در پیش بینی عملکرد گندم در استان خراسان رضوی به تفکیک شهرستان ها

Table 5 – Four evaluation metrics of the Random Forest and XGBoost algorithms in predicting wheat yield in Khorasan Razavi province, broken down by county

County	R^2 (XGB)	R^2 (RF)	RMSE (XGB)	RMSE (RF)	MAE (XGB)	MAE (RF)	d (XGB)	d (RF)
Torbat-E-Jam	0.13	0.34	269.36	312.61	211.48	308.78	0.77	0.71
Neyshabur	0	0.33	377.91	343.82	298.37	232.39	0.78	0.59

Iranian Journal of Field Crops Research

County	R^2 (XGB)	R^2 (RF)	RMSE (XGB)	RMSE (RF)	MAE (XGB)	MAE (RF)	d (XGB)	d (RF)
Mahvelat	0	0.16	481	344.22	441.6	329.94	0.71	0.69
Fariman	0	0.07	205.41	314.79	175.52	253.8	0.62	0.75
Bardeskan	0	0.05	387.03	187.69	341.11	149.23	0.6	0.8
Chenaran	0	0.05	410.79	281.64	317.07	202.42	0.59	0.62
Kashmar	0	0	472.61	217.93	424.28	164.5	0.51	0.56
Gonabad	0	0	683.93	444.72	579.88	372.6	0.46	0.55
Taybad	0	0	455.4	521.03	380.93	473.99	0.44	0.52
Mashhad	0	0	223.05	454	167.03	350.58	0.43	0.51
Torbat-E	0	0	786.2	581.33	685.32	505.85	0.39	0.47
Daregaz	0	0	678.69	432.65	484.36	388.38	0.38	0.44
Sabzevar	0	0	465.75	345.44	311.6	234.95	0.35	0.44
Khalil Abad	0	0	247.96	141.19	221.54	125.68	0.33	0.37
Quchan	0	0	943.99	667.65	797.74	622.39	0.25	0.36
Roshtkhar	0	0	515.13	352.85	415.8	285.01	0.2	0.18
Sarakhs	0	0	663.58	622.78	551.74	532.25	0.08	0.11
Kalat-E-Nader	0	0	464.9	609.82	413.58	450.16	0.06	0.1
Khaf	0	0	619.28	331.28	534.82	315.66	0.02	0.1
Mean	0.0068	0.0526	492.21	395.13	408.09	331.5	0.4194	0.4668

توضیح: مقایسه هر دو مدل در شهرستان‌ها بر اساس معیارهای ارزیابی مقدار R^2 و d مدل رندوم فارست انجام گرفته است (مقادیر شاخص ها صعودی و ترتیب اهمیت شهرستان ها نزولی مرتب شده است).

Note: The comparison of both models in the counties is based on the evaluation metrics of R^2 and d for the Random Forest model (the index values are ordered in ascending order, and the counties are arranged in descending order of importance)

- **ضریب تبیین:** این مقدار نشان‌دهنده میزان تطابق پیش‌بینی‌ها با داده‌های واقعی است و هرچه این مقدار به یک نزدیک‌تر باشد، مدل بهتر عمل کرده است. درصد تفاوت در R^2 نشان‌دهنده این است که مدل‌ها چقدر در تطبیق داده‌ها موفق بوده‌اند. مدل رندوم فارست در شش شهرستان مقدار ضریب تبیین بالاتری دارد. این موضوع نشان می‌دهد که این مدل در بیشتر موارد توانسته است توضیح‌دهنده بهتری برای داده‌ها باشد. وجود مقادیر صفر در ستون های ضریب تبیین، در نتیجه نرمال سازی این مقادیر ضریب تبیین حاصل شده است. نرمال سازی ضرایب تبیین در برخی موارد خاص مثل برخی مدل‌های رگرسیون غیرخطی یا رندوم فارست، امری معمول است.
- **میانگین قدرمطالق خطا:** نشان‌دهنده خطای مطلق پیش‌بینی است، هرچه کمتر باشد، مدل بهتر است. برای MAE ، در پانزده شهرستان‌ها مدل رندوم فارست عملکرد بهتری دارد، که نشان‌دهنده این است که این مدل توانسته است پیش‌بینی‌های دقیق‌تری انجام دهد.
- **شاخص توافق ویلموت** نیز برای اندازه‌گیری تطابق پیش‌بینی‌ها با مقادیر واقعی استفاده می‌شود و هرچه به ۱ نزدیک‌تر باشد، مدل بهتر است. شاخص d برای ارزیابی دقت مدل در شبیه‌سازی داده‌ها مفید است. رندوم فارست در پانزده شهرستان شامل بردسکن، سرخس، تربت جام، رشتخوار، نیشابور، مه ولات، کاشمر و گناباد امتیاز بالاتری دارد، که نشان‌دهنده دقت بیشتر در پیش‌بینی است.

Iranian Journal of
Field Crops Research

تعیین الگوریتم برتر در هر شهرستان بر اساس معیارهای ارزیابی و رتبه بندی شهرستان ها: مدل بهتر برای هر شهرستان با استفاده از میانگین چهار معیار ارزیابی هر مدل تعیین شد (جدول ۶). این مقایسه نشان می‌دهد که کدام مدل در پیش‌بینی دقیق‌تر عملکرد گندم در هر شهرستان عمل کرده است. طبق تحلیل انجام‌شده، مدل رندوم‌فارست به عنوان مدل برتر برای اکثر شهرستان‌ها شناخته شده است. این موضوع به وضوح در شهرستان‌هایی که رندوم‌فارست امتیاز بهتری نسبت به ایکس‌جی‌بوست کسب کرده است، مشاهده می‌شود. مدل ایکس‌جی‌بوست در چند شهرستان خاص، مانند بردسکن، مشهد، درگز و سرخس، امتیاز بهتری نسبت به رندوم‌فارست داشته است. این تفاوت‌ها ممکن است به ویژگی‌های خاص داده‌های این شهرستان‌ها، مانند توزیع متفاوت داده‌ها یا پیچیدگی‌های خاص این مناطق، برگردد. برای شش شهرستان شامل کاشمر، چناران، خاف، مه ولات، کلات نادر، رتبه ها در هر دو مدل مشابه بود (۳۱/۵ درصد کل شهرستان ها) (جدول ۶). در هفت شهرستان شامل خایا آباد، سبزوار، تربت جام، رشتخوار، نیشابور، گناباد و تربت حیدریه، مدل رندوم فارست امتیازهای بیشتری نسبت به مدل ایکس جی بوست داشت. شهرستان های رتبه های اول تا پنجم بر اساس میانگین معیارهای ارزیابی دو الگوریتم، همان شهرستان های رتبه های یک تا پنج در هر دو مدل بود (جدول ۶).

متوسطهای جذرمیانگین مربعات خطا و میانگین قدرمطالق خطا در مدل برنده عموماً کمتر است. برای مثال در شهرستان بردسکن، جذر میانگین مربعات خطا در مدل رندوم‌فارست برابر با ۱۸۷/۶ و در ایکس جی بوست برابر با ۲۰۵/۴ است که به‌وضوح اختلاف را نشان می‌دهد. در شهرستان‌هایی که ایکس‌جی‌بوست برتر بوده (مثل مشهد، رتبه ششم در برابر رتبه چهاردهم در رندوم فارست، مقدارهای آر ام اس ای تقریباً برابر و میانگین قدرمطالق خطا در مدل ایکس جی بوست اندکی کمتر است (۴۱۰/۷ در برابر ۴۵۴)، که حاکی از پیش‌بینی کمتر خطاتوسط ایکس‌جی‌بوست در داده‌های آن شهرستان است. شاخص توافق ویلموت که شباهت پیش‌بینی به مشاهدات را می‌سنجد، در مدل برنده (رندوم فارست) بالاتر است. در مه ولات مقدار d برای رندوم فارست رندوم‌فارست برابر ۰/۶۹ و برای ایکس‌جی‌بوست برابر ۰/۵۱ است؛ این تفاوت نشان می‌دهدکه رندوم فارست در آن ناحیه خطای پیش‌بینی کمتر داشته و الگوهای مشاهداتی را بهتر بازتولید می‌کند. در مناطقی که پراکندگی داده زیاد است (نقاطی با جذرمیانگین مربعات خطای بالا، شامل همه شهرستان ها به جز خاف، مشهد، تایباد، و سرخس) رندوم فارست معمولاً پایدارتر است و خطای کمتری ثبت می‌کند. ایکس‌جی‌بوست در مناطقی با توزیع پیچیده‌تر و نویز کمتر) مثل مشهد، درگز، تایباد، و سرخس توانسته کمی بهتر یا برابر با رندوم‌فارست عمل کند. این تحلیل نشان می‌دهد که علی‌رغم پتانسیل بالای ایکس‌جی‌بوست در موارد خاص، مدل رندوم فارست در اغلب شهرستان‌ها نتایج بهتری از نظر دقت پیش‌بینی (R^2) ، کاهش خطا (RMSE/MAE) و شباهت به مشاهدات (Willmott's d) ارائه کرده است. این تفاوت در عملکرد ممکن است ناشی از نحوه یادگیری و ویژگی‌های مدل‌ها باشد که در برخی موارد، رندوم‌فارست قادر است روابط پیچیده‌تری را شبیه‌سازی کند (Khodjaev et al., 2025). به‌منظور بهبود کلی، توصیه می‌شود بر روش‌های ترکیبی[1] یا تنظیم دقیق پارامترها[2] تمرکز شود تا بیش برازش در ایکس‌جی‌بوست و کم برازش احتمالی در رندوم فارست کاهش یابد.

جدول ٦- رتبه بندی شهرستان ها از نظر دقت و توان مدل بهینه شده بر اساس معیارهای ارزیابی و تعیین اعتبار در دو مدل رندوم فارست، ایکس جی بوست و میانگین آنها (رتبه در سمت چپ شهرستان آورده شده)

Table 6 – Ranking of counties based on the accuracy and performance of the optimized model according to evaluation metrics and validation in the two models, Random Forest, XGBoost, and their average (rank listed on the left side of the county)

[1] Ensemble
[2] Tuning

Avg Rank Counties	XGB Counties	RF Rank Counties
1 Khalil Abad	2 Khalil Abad	1 Khalil Abad
2 Bardeskan	1 Bardeskan	2 Bardeskan
3 Kashmar	3 Kashmar	3 Kashmar
4 Chenaran	4 Chenaran	4 Chenaran
5 Fariman	5 Fariman	5 Fariman
6 Sabzevar	8 Sabzevar	7 Sabzevar
7 Khaf	10 Khaf	10 Khaf
8 Mashhad	6 Mashhad	12 Mashhad
9 Torbat-E-Jam	12 Torbat-E-Jam	8 Torbat-E-Jam
10 Roshtkhar	13 Roshtkhar	9 Roshtkhar
11 Mahvelat	11 Mahvelat	11 Mahvelat
12 Daregaz	9 Daregaz	14 Daregaz
13 Tavbad	7 Tavbad	15 Tavbad
14 Nevshabur	15 Nevshabur	6 Nevshabur
15 Gonabad	17 Gonabad	13 Gonabad
16 Kalat-E-Nader	16 Kalat-E-Nader	16 Kalat-E-Nader
17 Sarakhs	14 Sarakhs	18 Sarakhs
18 Ouchan	18 Ouchan	19 Ouchan
19 Torbat-E Heydariveh	19 Torbat-E Heydariveh	17 Torbat-E Heydariveh

مقادیر ضریب تبیین برابر با صفر، به‌طور طبیعی به دلیل ضعف مدل در پیش‌بینی داده‌هاست. سورت کردن بر اساس ضریب تبیین و شاخص توافق کمک می‌کند که مدل‌هایی که به ترتیب بیشترین دقت پیش‌بینی را دارند و عملکرد خوبی در مقایسه با داده‌های واقعی دارند، در صدر قرار بگیرند. به منظور تصمیم گیری بهتر در خصوص استفاده از نوع مدل در هر شهرستان، در ستون آخر زیر عنوان مدل برتر، شهرستان ها بر اساس میانگین هر چهار معیار ارزیابی در هر دو الگوریتم، مرتب شدند. با این روش، با استفاده از میانگین مدل‌های دقیق‌تر و معتبرتر برای هر شهرستان، شهرستان ها از بهترین تا ضعیف ترین در ستونب مجزا در جدول ۶ قرار گرفتند.

مقایسه و ارزیابی مدل‌های یادگیری ماشین در طبقه‌بندی عملکرد گندم با استفاده از معیارهای دقت و بازیابی

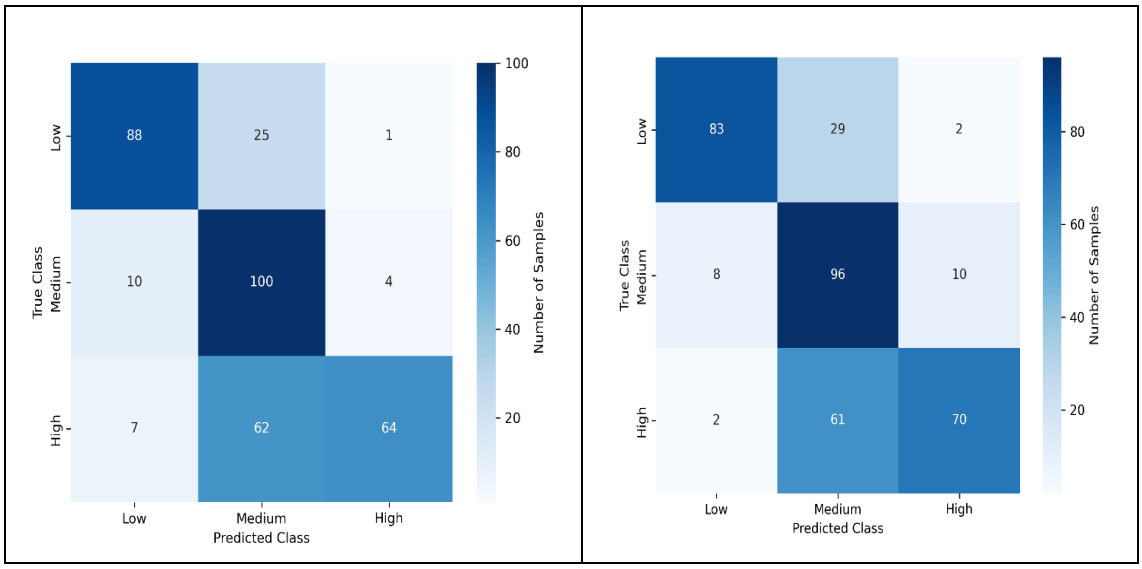
ابتدا داده‌های خام هر شهرستان به تفکیک سال خوانده شده و مدل‌های ذخیره‌شده با استفاده از کتابخانه جولیب[1] بارگذاری شدند. داده‌های آب‌وهوایی، اقلیمی و اکوفیزیولوژیک به عنوان ورودی مدل‌ها استفاده گردید و پس از آموزش، عملکرد مدل‌ها با شاخص‌های مختلف ارزیابی شد. در این مطالعه، دو الگوریتم یادگیری ماشین شامل رندوم‌فارست و ایکس‌جی‌بوست برای پیش‌بینی عملکرد گندم در سطح شهرستان‌ها توسعه داده شد. مقدار عملکرد ابتدا به‌صورت عددی با استفاده از رگرسیون پیش‌بینی گردید و سپس با استفاده از صدک‌های آماری به سه کلاس مجزا تقسیم شد:

- ضعیف (Low): صدک ۰ تا ۳۳؛ متوسط (Medium): صدک ۳۳ تا ۶۶؛ خوب (High): صدک ۶۶ تا ۱۰۰

[1] joblib

Iranian Journal of Field Crops Research

این طبقه‌بندی بر اساس مقادیر واقعی و همچنین مقادیر پیش‌بینی‌شده برای هر شهرستان انجام شده است. نتایج این طبقه‌بندی در قالب ماتریس درهم‌ریختگی (confusion matrix) برای هر مدل ارائه گردیده است (شکل ۸).



شکل ۸- ماتریس درهم‌ریختگی الگوریتم رندوم‌فارست (سمت راست) و ایکس‌جی‌بوست (سمت چپ) در طبقه بندی عملکرد گندم در استان خراسان رضوی

Figure 8 – Confusion matrix of the Random Forest algorithm (right) and XGBoost algorithm (left) in classifying wheat yield in Khorasan Razavi province

در ماتریس ایکس‌جی‌بوست، مقادیر قطر اصلی (مقدارهای صحیح) بسیار بیشتر هستند و در نتیجه ماتریس متراکم‌تر روی قطر است. در ماتریس رندوم‌فارست، خطاهای بیشتری در خارج از قطر اصلی دیده می‌شود که به معنی طبقه‌بندی اشتباه بالاتر است.

مقایسه این دو مدل بر اساس امتیاز کلی مدل ها که نشان‌دهنده عملکرد کلی هر مدل در پیش‌بینی عملکرد گندم برای شهرستان‌های مختلف است، صورت می گیرد. این امتیاز به‌طور خاص به شبیه‌سازی عملکرد مدل در مجموع شهرستان‌ها توجه دارد و بر اساس معیارهای مشخصی مانند دقت کلی، دقت مثبت یا طبقه‌ای، بازیابی، و امتیاز اف وان مدل در پیش‌بینی شهرستان‌های خاص محاسبه می‌شود (جدول ۷).

جدول ۷- خلاصه دقت(Accuracy) ، دقت مثبت(Precision) ، بازیابی (Recall) و امتیاز اف-وان برای دو الگوریتم طبقه بندی عملکرد گندم در استان خراسان رضوی

Table 7 – Summary of Accuracy, Precision, Recall, and F1 score for the two algorithms in classifying wheat yield in Khorasan Razavi province

مدل Model	دقت (%Accuracy)	دقت مثبت (%Precision)	بازیابی (%Recall)	F1-Score (%)
رندوم‌فارست Random Forest	68.9	78.0	68.9	67.7
ایکس‌جی‌بوست XGBoost	69.8	73.8	69.8	66.8

اعتبار سنجی: هر دو مدل ایکس‌جی‌بوست و رندوم‌فارست در پیش‌بینی عملکرد شهرستان‌ها قابل اعتماد هستند، با این تفاوت که ایکس‌جی‌بوست به طور کلی کلاس های عملکرد را با دقت بالاتری تفکیک کرده است. با این حال، مدل ایکس‌جی‌بوست

به‌طور کلی دقت بالاتری را نسبت به رندوم‌فارست ارائه داده است (۶۹/۸ درصد در برابر ۶۸/۹ درصد)، که نشان‌دهنده توانایی بیشتر این مدل در شناسایی صحیح کلاس‌های عملکرد می‌باشد. این نتیجه منطقی به نظر می‌رسد، زیرا ایکس‌جی‌بوست با بهره‌گیری از تکنیک‌های بوستینگ، توانایی بیشتری در مدل‌سازی الگوهای پیچیده و اصلاح خطاهای مدل‌های ضعیف‌تر دارد.

مقایسه شاخص‌های کلیدی نیز تفاوت‌هایی را بین این دو مدل نشان می‌دهد:

مقایسه دقیق‌تری از معیارهای دقت، دقت طبقه‌ای، و بازیابی برای هر مدل می‌تواند به درک بهتر نقاط قوت و ضعف هر مدل کمک کند. در این تحقیق، عملکرد بیشتر شهرستان‌ها با ایکس‌جی‌بوست پیش‌بینی شده‌اند، اما در برخی موارد خاص، رندوم فارست عملکرد بهتری نشان داده است که می‌تواند به نوع داده‌ها و ویژگی‌های آن‌ها بستگی داشته باشد.

۱. **دقت یا صحت** ایکس‌جی‌بوست با دقت کلی ۶۹/۸ درصد عملکرد بهتری نسبت به رندوم‌فارست (۶۸/۹ درصد) دارد. فقط درصد درست پیش‌بینی شده‌ها را می‌سنجد و اگر داده‌ها نامتوازن باشند (مثلاً کلاس‌های زیاد با هم فرق داشته باشند)، می‌تواند گمراه‌کننده باشد.

۲. **دقت مثبت یا دقت طبقه ای** برخلاف دقت کلی، در این شاخص، رندوم‌فارست عملکرد بهتری از خود نشان داده است ۷۸ درصد در برابر ۷۳/۸ درصد). این بدان معناست که رندوم‌فارست نسبت به ایکس‌جی‌بوست در کاهش خطاهای مثبت کاذب[1] موفق‌تر بوده است. دقت مثبت، دقت پیش‌بینی کلاس مثبت را نشان می‌دهد (چند درصد پیش‌بینی‌های مثبت درست بودند).

۳. **بازیابی** درصد نمونه‌های مثبت واقعی که مدل به درستی شناسایی کرده است. این معیار نشان می‌دهد که مدل چه مقدار از کل نمونه‌های مثبت را شناسایی کرده است. به عبارت دیگر، بازیابی به طور خاص به درصدی از نمونه‌های مثبت واقعی که به درستی توسط مدل شناسایی شده‌اند، اشاره دارد. به عبارت ساده‌تر، بازیابی به مدل می‌گوید که از تمامی نمونه‌های که باید به کلاس مثبت تعلق داشته باشند، چه مقدار توسط مدل به درستی شناسایی شده‌اند.

فرمول بازیابی به این صورت است: $Recall = TP / (TP + FN)$

که در آن TP[2] نشان دهنده تعداد نمونه‌هایی که مدل به درستی به عنوان مثبت شناسایی کرده است.

FN[3] تعداد نمونه‌هایی که مدل به اشتباه به عنوان منفی شناسایی کرده است، در حالی که باید مثبت می‌بودند. در نتیجه، بازیابی بالا به معنای این است که مدل به خوبی قادر است نمونه‌های مثبت را شناسایی کند، اما ممکن است همراه با اشتباهاتی (FN) باشد. به این ترتیب، بازیابی می‌تواند در برخی موارد از دقت بالاتر اهمیت بیشتری پیدا کند، به خصوص زمانی که اشتباه در شناسایی نمونه‌های مثبت پیامدهای جدی‌تری داشته باشد.

هر دو مدل در بازیابی کلاس‌های صحیح عملکرد مشابهی دارند، اما ایکس‌جی‌بوست با مقدار ۶۹/۸ درصد اندکی بهتر از رندوم‌فارست (۶۸/۹ درصد) عمل کرده است. این شاخص نشان‌دهنده توانایی مدل در شناسایی نمونه‌های واقعی هر کلاس است.

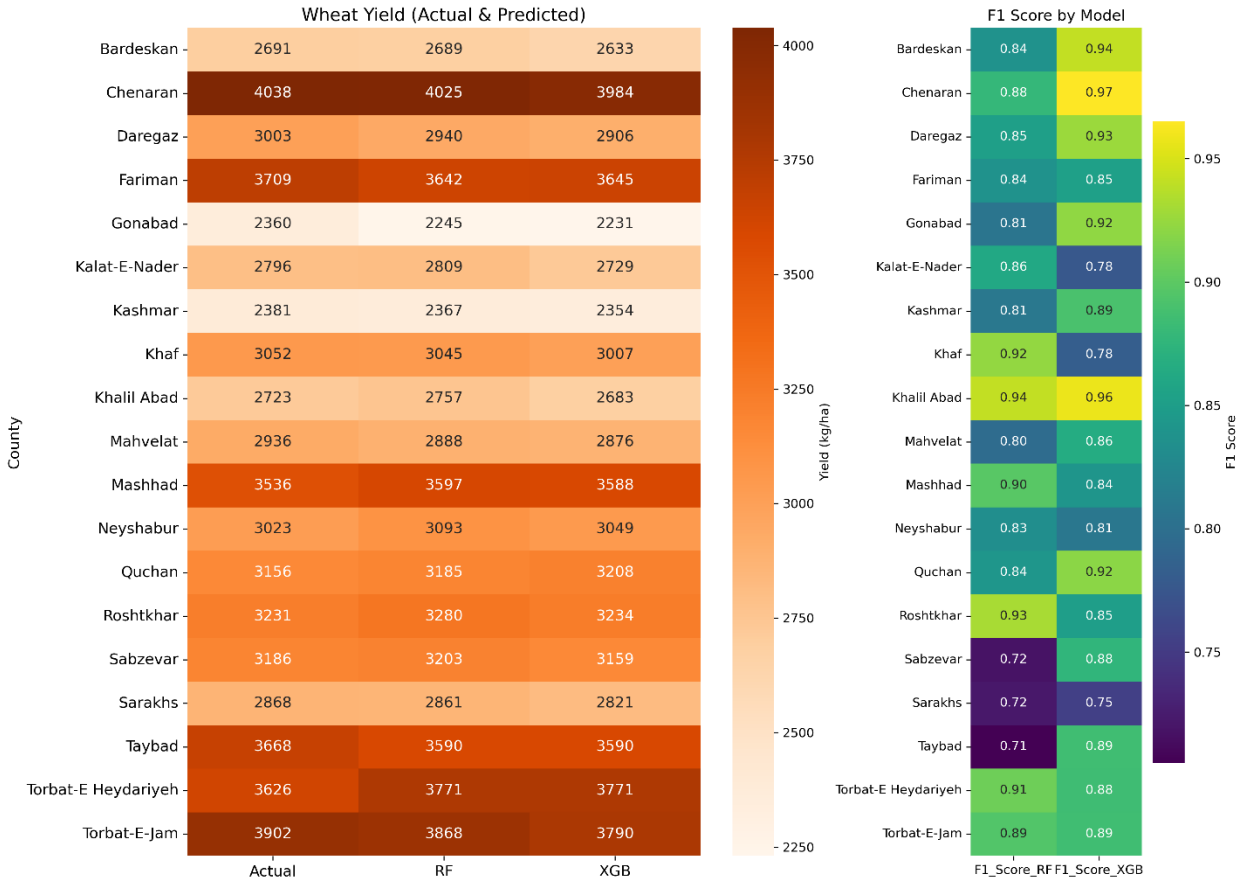
٤. **امتیاز اف-وان،** ترکیبی از دقت مثبت و بازیابی است که میانگینی متوازن از این دو معیار را می‌دهد و در مسائل کلاس‌بندی مفید واقع می شود. رندوم فارست با مقدار ۶۷/۷ درصد اندکی بهتر از ایکس جی بوست (۶۶/۸ درصد) بود، که بیانگر تعادل بهتر بین دقت و بازیابی در این مدل است. بهترین شاخص است وقتی که داده‌ها نامتوازن باشند یا هم خطاهای مثبت و هم منفی اهمیت داشته باشند.

[1] False Positives
[2] True Positives
[3] False Negatives

به طور کلی، اگرچه ایکس‌جی‌بوست از نظر دقت کلی کمی برتر است، اما رندوم‌فارست در دقت طبقه‌ای وامتیاز اف-وان عملکرد قابل‌توجهی دارد. این تفاوت‌ها می‌توانند ناشی از ویژگی‌های داده‌ها، نوع توزیع کلاس‌ها، یا حساسیت مدل‌ها نسبت به ویژگی‌های خاص باشند. بنابراین انتخاب نهایی مدل باید با در نظر گرفتن هدف هایی مدل، نوع کاربرد و پیامدهای هر نوع خطا انجام شود. به عنوان مثال، در شرایطی که هدف کاهش هشدارهای اشتباه یا پیش‌بینی‌های نادرست مثبت است، مدلی با دقت بالاتر مانند رندوم‌فارست ترجیح داده می‌شود. اما اگر هدف شناسایی کامل‌تر تمامی موارد واقعی باشد، مدلی با بازیابی بالاتر مانند ایکس‌جی‌بوست می‌تواند مفیدتر باشد. افزون بر این، میزان پیچیدگی مدل، زمان آموزش، و نیاز به تنظیم پارامترها نیز باید در تصمیم‌گیری لحاظ شود. رندوم‌فارست عموماً ساختار ساده‌تری دارد و به تنظیمات کمتری نیاز دارد، در حالی که ایکس‌جی‌بوست گرچه پیچیده‌تر و زمان‌برتر است، اما پتانسیل بیشتری برای استخراج الگوهای دقیق‌تر در داده‌های پیچیده دارد. در نهایت، ارزیابی عملکرد مدل‌ها باید همیشه در بستر کاربرد واقعی و با توجه به نیازهای تصمیم‌گیری انجام گیرد، چرا که یک مدل از نظر آماری بهتر ممکن است در عمل کاربردپذیری کمتری داشته باشد.

در شکل۹، اف وان-اسکور دو مدل برای تمام شهرستان ها نشان داده شده، تا تحلیل عمیق تر را ممکن سازد. مطابق این شکل، مدل رندوم فارست در پیش بینی عملکرد گندم شهرستان های بردسکن، چناران، درگز، گناباد، کلات نادر، کاشمر، خاف، خلیل آباد، مه ولات، مشهد، نیشابور، رشتخوار، سبزوار، سرخس، و تربت جام بهتر از ایکس‌جی‌بوست عمل کرده است. مدل ایکس‌جی‌بوست در پیش بینی عملکرد گندم تنها دو شهرستان فریمان (با اختلاف سه کیلوگرم)، و قوچان بهتر از رندم فارست عمل کرد. برای د.و شهرستان تربت حیدریه و تایباد پیش بین دو مدل عیناً یکسان بود. مقایسه مقادیر اف-وان اسکور، (شکل ۹) نشان می دهد که این مقادیر برای هر دو مدل بسیار نزدیک بوده اند، و برای مدل ایکس‌جی‌بوست اندکی بیشتر بود ه است، به عبارت دیگر پیش بینی های این مدل به مقادیر واقعی عملکرد نزدیک تر بودند.

Iranian Journal of Field Crops Research
Ferdowsi University of Mashhad



شکل ۹– مقادیر عملکرد گندم واقعی و پیش بینی شده توسط رندوم فارست و ایکس جی بوست در استان خراسان رضوی به همراه امتیاز اف–وان برای هر مدل و شهرستان

Figure 9 – Actual and predicted wheat yield values by Random Forest and XGBoost in Khorasan Razavi province, along with the F1 score for each model and county

تحلیل خطا و محدودیت‌های دو الگوریتم در طبقه بندی عملکرد گندم

در هر دو مدل، برچسب‌های واقعی و پیش‌بینی‌شده در سه‌کلاس پایین، متوسط و بالا طبقه‌بندی شده‌اند. نتایج بیشتر به منظور مقایسه دو مدل در جدول ۸ آورده شده است. در رندوم‌فارست، بیشترین اشتباهات مربوط به جابه‌جایی از «High» به «Medium»و از «Low» به «Medium» است و تنها دو مرتبه برچسب «High» مستقیماً به «Low» یا برعکس تغییر یافت که نشان دهندهٔ پراکندگی نسبتاً زیاد در مرزهای مجاور کلاس‌ها است. خطاهای طبقه‌بندی به نسبت زیاد دیده می‌شود. به‌ویژه در برخی سال‌ها، پیش‌بینی از Medium به Low یا از High به Medium تغییر کرده است. پراکندگی بین کلاس‌های پیش‌بینی شده نسبت به واقعیت زیاد است.

ایکس‌جی‌بوست با بهبود جزئی در دقت کلی و کاهش چشمگیر اشتباهات «Medium→High» و«Low→High» ، توانسته مرزهای میانی را دقیق‌تر تشخیص دهد؛ ولی در عوض اندکی اشتباهات شدید (High→Low) را افزایش داده است. به‌طور خاص، در مدل رندوم‌فارست موارد متعددی وجود داشت که شهرستان‌های با عملکرد ضعیف به‌اشتباه در کلاس متوسط یا حتی خوب قرار گرفتند. این موضوع در ایکس‌جی‌بوست با شدت کمتری مشاهده شد و نشان‌دهنده‌ی قدرت بیشتر این مدل در یادگیری الگوهای پیچیده‌تر از داده‌هاست. مدل رندوم‌فارست مناسب شرایطی است که داده‌ها کمی نویزدار یا پیچیده

باشـــند، چون درخت‌های تصـــادفی می‌توانند روی ویژگی‌های متفاوت تمرکز کنند. مدل ایکس‌جی‌بوســـت به خاطر مکانیزم بوستینگ، توانسته بهتر الگوهای دقیق‌تر را یاد بگیرد و رکورد بهتری ثبت کند.

جدول ۸- تحلیل خطاهای دو الگوریتم رندوم فارست و ایکس جی بوست در پیش بینی طبقه عملکرد گندم در استان خراسان رضوی

Table 8 – Error analysis of the two algorithms, Random Forest and XGBoost, in predicting wheat yield classification in Khorasan Razavi province

معیار / الگوی اشتباه	رندوم‌فارست	ایکس‌جی‌بوست
دقت کلی (Accuracy)	٪ 68.98	٪ 69.81
میانگین F1 Score	0.678	0.668
میانگین Precision	0.780	0.738
میانگین Recall	0.690	0.698
تعداد اشتباه High → Medium Error	61	62
تعداد اشتباه Low → Medium Error	29	25
تعداد اشتباه Medium → High Error	10	4
تعداد اشتباه Medium → Low Error	8	10
تعداد اشتباه High → Low Error	2	7
تعداد اشتباه Low → High Error	2	1

نکات مهم و پیشـنهاد: ایکس‌جی‌بوسـت: کاهش قابل‌توجه «Medium→High» و «Low→High»؛ اندکی دقت کلی بالاتر؛ افزایش اشتباهات شدید«High→Low» . رندوم‌فارست: امتیاز اف-وان و دقت مثبت بالاتر؛ اشتباهات شدید کمتر؛ پراکندگی بیشـــتر در اشـــتباهات بین کلاس‌های مجاور. اگر اولویت، تفکیک دقیق مرزهای میانی (کاهش خطاهای جابه‌جایی مجاور) و حداکثر دقت کلی است، ایکس‌جی‌بوست مناسب‌تر است. اگر اولویت، دستیابی به امتیاز اف-وان و دقت مثبت بالاتر و کاهش اشتباهات شدید باشد، رندوم‌فارست گزینه‌ی بهتری است. نتایج حاکی از آن است که در طبقه بندی کلا ســـهای عملکرد گندم، مدل رندوم‌فارســـت با مدل ایکس‌جی‌بوســـت تفاوت محســـوســـی ندارد، بنابراین بهبود عملکرد ایکس‌جی‌بوســـت به دلیل خود مدل اســـت نه تغییر در توزیع داده‌ها. به طور کلی، اگر هدف طبقه‌بندی بهتر عملکرد گندم آبی در شـــرایط اقلیمی اســـتان خراســـان رضـــوی به ســـه طبقه (کلاس) باشـــد، ایکس‌جی‌بوســـت انتخاب بســـیار بهتری نســـبت به رندوم‌فارست است و در صورت وجود منابع محاسباتی کافی، به کارگیری آن توصیه می شود.

۳- تحلیل عملکرد مدل‌ها بر اساس خوشه‌های مکانی شهرستان‌ها

برای بررسی دقیق‌تر عملکرد مدل‌های یادگیری ماشین در مناطق مختلف، شهرستان‌ها بر اساس شباهت‌های مکانی و اقلیمی به سه خوشه تقسیم شدند (شکل ۱۰). این خوشه‌بندی به کمک الگوریتم‌های تحلیل خوشه‌ای و بر پایه‌ی ویژگی‌های مکانی انجام شده است (مطالعات اقلیمی با استفاده از نقشه‌های جهانی به‌روزشده کوپن–گایگر (Peel et al., 2007) بر این تفکیک اقلیمی تأکید دارند):

- خوشه‌ی اول (جنوب و شرق استان؛ اقلیم بسیار خشک با تغییرات دمایی زیاد و دوره رشد بسیار گرم و طولانی) شهرستان های بردسکن، خلیل‌آباد، مه‌ولات و سرخس قرار دارند که در همه‌ی این شهرستان‌ها مدل رندوم‌فارست نسبت به ایکس‌جی‌بوست عملکرد بهتری ارائه داده است.
- خوشه‌ی دوم (مرکز استان؛ اقلیم نیمه‌خشک تا نیمه‌مرطوب با تغییرات دمایی کم و دوره رشد خنک‌تر و کوتاه‌تر) شامل شش شهرستان چناران، فریمان، کلات‌نادر، مشهد، نیشابور و قوچان است که در چهار شهرستانِ چناران، کلات‌نادر، مشهد و نیشابور، رندوم‌فارست عملکرد برتر داشته است. در دو شهرستانِ فریمان و قوچان، ایکس‌جی‌بوست با اختلاف اندک (حدود ۳ کیلوگرم در فریمان) بهتر عمل کرده است.
- خوشه‌ی سوم (غرب و جنوب غرب استان؛ اقلیم نیمه‌خشک گرم با تغییرات دمایی متوسط و دوره رشد نسبتاً گرم و طولانی) شامل ۹ شهرستان: درگز، گناباد، کاشمر، خاف، رشتخوار، سبزوار، تایباد، تربت‌جام و تربت‌حیدریه است که در در هفت شهرستان (درگز، گناباد، کاشمر، خاف، رشتخوار، سبزوار و تربت‌جام)، رندوم‌فارست دقت بالاتری نسبت به ایکس‌جی‌بوست نشان داد. در دو شهرستانِ تایباد و تربت‌حیدریه، هر دو مدل دقیقاً نتایج یکسانی ارائه داده‌اند.

با توجه به نتایج خوشه‌بندی ناحیه های اقلیمی (سه خوشه) و ادغام آن با کلاس‌بندی عملکرد دو مدل در پیش‌بینی عملکرد گندم، و همچنین توجه به این نکته که که مدل رندوم فارست در ۸۰ درصد کل شهرستان ها پیش بینی دقیق تری از مدل ایکس‌جی‌بوست ارائه داد، این مدل به عنوان گزینه‌ی مناسب‌تر برای تحلیل‌های آتی در حوزه‌ی پیش‌بینی عملکرد گندم بر مبنای ناحیه اقلیمی پیشنهاد می‌شود.

جدول ۹- مدل برتر در پیش بینی عملکرد گندم در استان خراسان رضوی بر مبنای خوشه بندی شهرستان ها در ناحیه های اقلیمی سه گانه

Table 9 – The best model for predicting wheat yield in Khorasan Razavi province based on the clustering of counties in the three climatic regions

City	Cluster	Better Model
Neyshabur	2	رندوم‌فارست RF
Chenaran	2	رندوم‌فارست RF
Quchan	2	ایکس‌جی‌بوست XGB
Gonabad	3	رندوم‌فارست RF
Torbat-e-Jam	3	رندوم‌فارست RF
Sabzevar	3	رندوم‌فارست RF
Roshtkhar	3	رندوم‌فارست RF
Daregaz	3	رندوم‌فارست RF
Bardeskan	1	رندوم‌فارست RF
Khalil Abad	1	رندوم‌فارست RF
Mahvelat	1	رندوم‌فارست RF
Sarakhs	1	رندوم‌فارست RF
Fariman	2	ایکس‌جی‌بوست XGB
Kalat-E-Nader	2	رندوم‌فارست RF
Mashhad	2	رندوم‌فارست RF
Torbat-E Heydarieh	3	Equal

City	Cluster	Better Model
Taybad	3	Equal
Khaf	3	رندوم‌فارست
Kashmar	3	رندوم‌فارست

۱ اقلیم بسیار خشک با تغییرات دمایی زیاد و دوره رشد بسیار گرم و طولانی.

۲ اقلیم نیمه‌خشک تا نیمه‌مرطوب با تغییرات دمایی کم و دوره رشد خنک‌تر و کوتاه‌تر.

۳ اقلیم نیمه‌خشک گرم با تغییرات دمایی متوسط و دوره رشد نسبتاً گرم و طولانی.

1 Hyper-arid climate with high temperature variations and a very hot and long growing season.
2 Semi-arid to semi-humid climate with low temperature variations and a cooler and shorter growing season.
3 Warm semi-arid climate with moderate temperature variations and a relatively warm and long growing season.

نتیجه گیری کلی

دو مدل پس از بهینه سازی با رگرسورهای مربوط به خود، تفاوت قابل توجهی از نظر توان و دقت پیش بینی نداشتند، با اینحال، با توجه به عملکرد اندکی دقیق‌تر رندوم فارست، این مدل می‌تواند گزینه‌ای مناسب‌تر برای استفاده در سیستم‌های تصمیم گیری در حوزه کشاورزی باشد، به‌ویژه در شرایطی که شناسایی شهرستان‌های با عملکرد بالا (برای سرمایه‌گذاری یا الگوبرداری) یا شهرستان‌های با عملکرد پایین (برای مداخلات حمایتی) اهمیت داشته باشد، مدل رندوم فارست می‌تواند تصمیم‌گیری های بهتری فراهم کند.

قدردانی

بخشی از هزینه های این تحقیق توسط معاونت محترم پژوهش و فناوری دانشگاه فردوسی مشهد، در قالب طرح پژوهشی به شماره ۲/۶۴۲۶۴ مصوب ۱۴۰۴/۰۲/۲۱ تأمین شده است که به این وسیله قدردانی می شود. از جناب آقای مهندس مهدی سلطانی، دانش‌آموخته کارشناسی ارشد اگرواکولوژی، که بخش عمده داده های خام مورد نیاز را در اختیار قرار دادند، سپاسگزارم.

References

1. Aslan, M. F., Sabanci, K., & Aslan, B. (2024). Artificial intelligence techniques in crop yield estimation based on Sentinel-2 data: A comprehensive survey. Sustainability, 16(18), 8277. RePEc:gam:jsusta:v:16:y:2024:i:18:p:8277-:d:1483971
2. Asseng, S., Ewert, F., Martre, P., Rötter, R. P., Lobell, D. B., Cammarano, D., et al. (2015). Rising temperatures reduce global wheat production. Nature Climate Change, 5(2), 143–147. https://doi.org/10.1038/nclimate2470
3. Belete, Y.S. 2024. Wheat Yield Prediction Based on Random Forest Method. Preprints.org (www.preprints.org) NOT PEER-REVIEWED Posted: 24 February 2025. https://doi.org/10.20944/preprints202502.1837.v1
4. Dhillon, M. S., Dahms, T., Kuebert-Flock, C., Rummler, T., Arnault, J., Steffan-Dewenter, I., & Ullmann, T. (2023). Integrating Random Forest and crop modeling improves the crop yield prediction of winter wheat and oil seed rape. Frontiers in Remote Sensing, 3, 1010978. https://doi.org/10.3389/frsen.2022.1010978
5. Everingham, Y., Sexton, J., Skocaj, D., Inman-Bamber, G. (2016) Accurateprediction of sugarcane yield using a random forest algorithm. Agronomy Sustain Dev 36(2): 1–9. https:// doi. org/ 10. 1007/s13593- 016- 0364-z

6. Farhadi, M., Bannayan, M., M.H. Fallah, Jahan, M. (2024). Identiication of climatic and management factors inluencing wheat's yield variability using AgMERRA dataset and DSSAT model across a temperate region. Discover Life, https://doi.org/10.1007/s11084-024-09651-8
7. Gheysarbeigi, S., Pir Bavaghar, M., Valipour, A. (1403). Forest Aboveground Biomass Estimation Using Satellite Imagery and Random Forest Regression Model. Geography and Environmental Sustainability, 14 (1): 85-100. DOI:10.22126/GES.2024.9971.2715
8. Hatfield, J. L., & Prueger, J. H. (2015). Temperature extremes: Effect on plant growth and development. Weather and Climate Extremes, 10, 4–10. https://doi.org/10.1016/j.wace.2015.08.001
9. IPCC, 2022. Sixth Assessment Report. Available Online at: https://www.ipcc.ch/site/assets/uploads/2022/04/AR6_Factsheet_April_2022.pdf
10. Jhajharia, K., Mathur, P., Jain, S., & Nijhawan, S. (2023). Crop yield prediction using machine learning and deep learning techniques. Procedia Computer Science, 218, 406-417. https://doi.org/10.1016/j.procs.2023.01.023
11. Khodabandehloo, E., Azadbakht, M., Radiom, S., Ashourloo, D., Alimohammadi, A. (1400). Prediction of Wheat Fusarium Head Blight Severity by Using Random Forest. Iranian Remote Sensing & GIS, 13(4). 1-44.
12. Khodjaev, S., Bobojonov, I., Kuhn, L., Glauben, T. (2025). Optimizing machine learning models for wheat yield estimation using a comprehensive UAV dataset. Modeling Earth Systems and Environment, 11:15. https://doi.org/10.1007/s40808-024-02188-9
13. Krishnadoss, N., Ramasamy, L.K. (2024) Crop yield prediction with environmental and chemical variables using optimized ensemble predictive model in machine learning. Environ Res Commun 6(10): 101001. https://doi.org/10.1088/2515-7620/ad7e81
14. Monavar Sabegh, S., Zare Haghi, D., Samadianfard, S., Neishabouri, M.R., Mikaeili, F. (1402). Estimation of Daily Reference Evapotranspiration Using Random Forest Optimized by Genetic Algorithm. Water and Soil Science, 33(4): 33-53. DOI: 10.22034/ws.2021.48756.2449
15. Nayak, H.S., Silva, J.V., Parihar, C.M., Krupnik, T.J., Sena, D.R., Kakraliya, S.K., Jat, H.S., Sidhu, H.S., Sharma, P.C., Jat, M.L., et al. (2022). Interpretable machine learning methods to explain on-farm yield variability of high productivity wheat in Northwest India. Field Crop. Res. 2022, 287, 108640. https://doi.org/10.1016/j.fcr.2022.108640
16. Oghnoum, M., Feghhi, J., Makhdoum, M., Moghaddamnia, A., Etemad, V. (1398). Land capability evaluation of afforestation using Random Forest algorithm (Kan Watershed, Tehran). Journal of Forest Research and Development, 5(3): 387-403.
17. Pang, A., Chang, M.W., Chen, Y. (2022). Evaluation of Random Forest for Regional and Local-Scale Wheat Yield Prediction in Southeast Australia. Sensors 2022, 22, 717. https://doi.org/10.3390/s22030717
18. Peel, M. C., Finlayson, B. L., McMahon, T. A. (2007). *Updated world map of the Köppen–Geiger climate classification.* Hydrology and Earth System Sciences, 11(5): 1633–1644. https://doi.org/10.5194/hess-11-1633-2007
19. Remman SB, Lekkas AM (2021) Robotic lever manipulation using hindsight experience replay and shapley additive explanations. 2021 European Control Conference (ECC). https://doi.org/10.23919/ecc54610.2021.9654850
20. Roell, Y. E., Beucher, A., Møller, P. G., Greve, M. B., & Greve, M. H. (2020). Comparing a Random Forest based prediction of winter wheat yield to historical yield potential. Agronomy, 10(3), 3. https://doi.org/10.3390/agronomy10030395
21. Sadeghi, M., Ahmadi Nadoushan, M. (1400). Modeling Soil Nitrogen Using Remote Sensing, Regression and Random Forest Models. Journal of Water and Soil Resources Conservation (WSRCJ), 11(2): 97-111. DOI: 10.30495/WSRCJ.2021.19216

22. Shahabi, S., Azarpira, F., Barzkar, A. (1399). Estimation of Daily and Weekly Evapotranspiration Using Hybrid Approaches of Soft Computing. Iranian Journal of Irrigation and Drainage, 5(14): 1550-1561.
23. Shen, Y., Mercatoris, B., Liu, Q., Yao, H., Li, Z., Chen, Z. (2024). Use Self-Training Random Forest for Predicting Winter Wheat Yield. Remote Sens., 16, 4723. https://doi.org/10.3390/rs16244723
24. Si, Z., Qin, A., Liang, Y., Duan, A., & Gao, Y. (2023). A Review on Regulation of Irrigation Management on Wheat Physiology, Grain Yield, and Quality. Plants, 12(4), 692. https://doi.org/10.3390/plants12040692
25. Vieira, H. V., Bradford, B. Z., Osterholzer, A., Peirce, E. S., Cockrell, D., Peairs, F., et al. (2025). A new growing degree-day phenology model for wheat stem sawfly (Hymenoptera: Cephidae) in Colorado wheat fields. PLoS ONE, 20(4), e0320497. https://doi.org/10.1371/journal.pone.0320497
26. Slafer, G. A., Savin, R., Sadras, V. O., & Calderini, D. F. (2023). Wheat yield improvement: physiological and agronomic basis. Field Crops Research, 291, 108757. https://doi.org/10.1016/j.fcr.2023.108757
27. Soleimannejad, L., Bonyad, A.E., Naghdi, R., Latifi, H. (2018). Classification of quantitative attributes of Zagros forest using Landsat 8-OLI and Random Forest algorithm (Case study: protected area of Manesht forests). Journal of Forest Research and Development, 4(4): 415-434.
28. Soltani, M., Jahan, M., Yaghoubi, F. (2024). Evaluation of power and accuracy of AgMERRA and ERA5 dataset to simulate long term data for temperature and radiation in grat Khorasan province. (Under referee process)
29. Taiz, L., E. Zeiger, I. M. Moller, et al. (2018). Fundamentals of plant physiology. New York, USA: Oxford University. Press. ISBN 978160535790
30. Ting, K.M. (2011). Confusion Matrix. In: Sammut, C., Webb, G.I. (eds) Encyclopedia of Machine Learning. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-30164-8_157
31. Yaghoubi, F., Bannayanm, M., Asadi, G. (2020). Performance of predicted evapotranspiration and yield of rainfed wheat in the northeast Iran using gridded AgMERRA weather data. International Journal of Biometeorology 64:1519–1537 https://doi.org/10.1007/s00484-020-01931-y
32. Wang, Z., Li, S. (2002). Effects of water deficit and supplemental irrigation at different growing stages on uptake and distribution of nitrogen, phosphorus, and potassium in winter wheat. Journal of Plant Nutrition and Fertilizers, 8(3), 265–270. https://doi.org/10.11674/zwyf.2002.0302