



Predicting Sugarcane Yield Using Sentinel-2 Vegetation Indices, K-Means Clustering, and K-Nearest Neighbors (KNN) Regression

Feryal Jaderi¹ | Nasim Monjezi²

1. Biosystems Engineering Department, Faculty of Agriculture, Shahid Chamran University of Ahvaz, Ahvaz, Iran, Email: feryaljaderi@gmail.com

2. Corresponding Author, Biosystems Engineering Department, Faculty of Agriculture, Shahid Chamran University of Ahvaz, Ahvaz, Iran, Email: n.monjezi@scu.ac.ir

Article Info	ABSTRACT
Article type: Research Article	Accurate prediction of the yield of strategic crops such as sugarcane plays a key role in optimal resource management and ensuring food security. The main objective of this study is to develop a robust and interpretable model based on the K-Nearest Neighbors (KNN) regression algorithm to predict sugarcane yield prior to harvest. To this end, Sentinel-2 satellite imagery was integrated with engineered agronomic features, including indicators of water and fertilizer use efficiency. In addition, the K-means clustering algorithm was employed to partition fields into homogeneous groups, enabling the KNN model to better capture spatial heterogeneity and improve prediction accuracy. Key vegetation indices were extracted from Sentinel-2 time-series imagery, and engineered features were generated to enrich the dataset. The proposed model achieved a coefficient of determination (R^2) of 0.8706 and a root mean square error (RMSE) of 7.80 t ha ⁻¹ on the test dataset. Feature importance analysis revealed that the engineered variables—particularly water productivity—were among the main predictors of yield. The results demonstrate that integrating satellite data with a simple yet effective KNN model provides a transparent and practical tool to support decision-making in precision agriculture.
Article history:	
Received: Sep. 27, 2025	
Revised: Oct. 18, 2025	
Accepted: Oct. 30, 2025	
Published online: Winter 2026	
Keywords: <i>K-Means clustering,</i> <i>KNN regression,</i> <i>Sentinel-2,</i> <i>Sugarcane,</i> <i>Yield prediction.</i>	
Cite this article Jaderil, F., Monjezi, N., (2025) Predicting Sugarcane Yield Using Sentinel-2 Vegetation Indices, K-Means Clustering, and K-Nearest Neighbors (KNN) Regression, <i>Iranian Journal of Biosystem Engineering</i> , 56 (4), 91-105. https://doi.org/10.22059/ijbse.2025.403177.665620	
© The Author(s). Publisher: The University of Tehran Press.	
DOI: https://doi.org/10.22059/ijbse.2025.403177.665620	

EXTENDED ABSTRACT

Introduction

Sugarcane is a globally important crop that serves as a primary source of sugar and a vital feedstock for biofuels. Traditional yield estimation methods are often labor-intensive, time-consuming, and limited in spatial coverage, highlighting the need for more scalable, data-driven approaches. The advent of remote sensing technologies has transformed agricultural monitoring by providing extensive, multi-temporal, and high-resolution datasets. While machine learning algorithms show promise, many existing models rely solely on spectral data and often overlook the predictive power of engineered agro-technical features that quantify farm management efficiency. The main objective of this study is to develop a robust and interpretable K-Nearest Neighbors (KNN) regression model for pre-harvest sugarcane yield prediction by integrating Sentinel-2 imagery with engineered agro-technical features. Key features, such as water and fertilizer use efficiency, are incorporated to enhance model performance. Additionally, K-means clustering is applied to group farms into homogeneous categories, enabling the KNN model to better capture spatial heterogeneity and improve prediction accuracy. Overall, the study aims to provide a reliable and transparent framework for estimating pre-harvest sugarcane yield using the synergy of satellite imagery and machine learning techniques.

Materials and Methods

This study was conducted on sugarcane farms managed by the Dehkhoda Sugarcane Agro-industry during the crop years 2017 to 2024. **Agronomic and management data:** The ground dataset consisted of records from 2,417 unique farm plots. Key variables included final gross yield (tons/ha), water consumption (m³/ha), total fertilizer application (kg/ha), and soil electrical conductivity (EC). Multi-temporal Sentinel-2 Level 2A (atmospherically corrected) images were used. Standard vegetation indices such as NDVI, EVI, and GNDVI were calculated from the Sentinel-2 bands. In addition, five engineered features were developed to quantify agro-technical efficiency and interactions: Water Use Efficiency, Fertilizer Efficiency, a Vegetation Health Index, the Irrigation-to-Fertilizer Ratio, and a Soil-Vegetation Interaction metric. The top 25 most informative features were then selected using the SelectKBest algorithm with an F-regression scoring function.

Two-step modeling approach:

1. The K-means algorithm was applied for unsupervised clustering of farm plots based on combined spectral and agro-technical features. The optimal number of clusters ($k=4$) was determined using the Elbow Method, and the resulting cluster labels were added as a new categorical feature to the dataset.

2. The K-Nearest Neighbors (KNN) algorithm was used to predict yield.

Model evaluation: The dataset was split into 80% training and 20% testing sets, and a 5-fold cross-validation technique was applied to the training data to ensure model robustness. Model performance was assessed using three standard metrics: the coefficient of determination (R^2), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE).

Results and Discussion

The statistical summary of the dataset revealed significant variability in final gross yield, ranging from 14.58 to 147.13 t/ha. The K-means clustering successfully partitioned the farm plots into four distinct agro-technical profiles, as confirmed by the mean feature values for each cluster. For example, Cluster 1 contained high-yielding plots with low soil salinity and high fertilizer use, while Cluster 2 included low-yielding plots with the lowest average fertilizer input. This spatial stratification was a crucial step, as it enabled the KNN model to learn from a more homogeneous set of neighboring plots, thereby enhancing predictive accuracy. The proposed KNN model demonstrated strong performance on the unseen test data, achieving an R^2 of 0.8706, RMSE of 7.80 t/ha, and MAE of 6.59 t/ha. The R^2 value indicates that the model explains approximately 87% of the variance in final gross yield. A scatter plot of predicted versus actual yields further confirmed the model's high accuracy, with data points closely aligned along the 1:1 line. Feature importance analysis revealed that the engineered features, particularly water use efficiency, were significant predictors of yield, highlighting the importance of integrating on-farm management indicators with remote sensing features.

Conclusion

This study successfully developed and validated a robust and highly interpretable framework for pre-harvest sugarcane yield prediction. By integrating unsupervised K-means clustering with KNN regression and enhancing it with engineered agro-technical features, the proposed model achieved an R^2 of 0.8706 on unseen test data. The findings demonstrate that a relatively simple, sample-based algorithm—when coupled with effective feature engineering and spatial stratification—can perform competitively with more complex black-

box models. The superior interpretability and computational efficiency of the KNN approach make it a practical and transparent tool for decision-making in real-world precision agriculture applications. Future work should focus on integrating dynamic meteorological data and exploring multi-task learning frameworks to predict multiple crop parameters simultaneously.

Author Contributions

Feryal Jaderi: Data collection and analysis, and initial writing

Nasim Monjezi: Study design, text revision, Providing expert opinions and reviews

Data Availability Statement

Not applicable

Acknowledgements

The authors would like to thank Shahid Chamran University of Ahvaz for providing funding for this research.

Ethical considerations

The authors avoided data fabrication, falsification, plagiarism, and misconduct.

Conflict of interest

The author declares no conflict of interest.

پیش‌بینی عملکرد نیشکر با استفاده از شاخص‌های پوشش‌های گیاهی ماهواره‌ای سنتینل ۲ و

الگوریتم‌های خوشه‌بندی K-MEANS و رگرسیون KNN

فریال جادری^۱ | نسیم منجزی^۲✉^۱ گروه مهندسی بیوسیستم، دانشکده کشاورزی، دانشگاه شهید چمران اهواز، اهواز، ایران، رایانامه: feryaljaderi@gmail.com^۲ نویسنده مسئول، گروه مهندسی بیوسیستم، دانشکده کشاورزی، دانشگاه شهید چمران اهواز، اهواز، ایران رایانامه: n.monjezi@scu.ac.ir

اطلاعات مقاله	چکیده
نوع مقاله: مقاله پژوهشی	
تاریخ دریافت: ۱۴۰۴/۷/۵	
تاریخ بازنگری: ۱۴۰۴/۷/۲۶	
تاریخ پذیرش: ۱۴۰۴/۸/۸	
تاریخ انتشار: زمستان ۱۴۰۴	
واژه‌های کلیدی:	پیش‌بینی دقیق عملکرد محصولات استراتژیک مانند نیشکر، نقشی کلیدی در مدیریت بهینه منابع و تضمین امنیت غذایی دارد. هدف اصلی این پژوهش، توسعه مدلی مقاوم و قابل تفسیر بر پایه الگوریتم رگرسیون K-نزدیک‌ترین همسایه (KNN) برای پیش‌بینی عملکرد نیشکر پیش از برداشت است. در این راستا، تصاویر ماهواره‌ای سنتینل ۲ با ویژگی‌های زراعی مهندسی‌شده، شامل شاخص‌های کارایی مصرف آب و کود، تلفیق شدند. افزون بر این، الگوریتم خوشه‌بندی K-میانگین برای تقسیم مزارع به گروه‌های همگن به کار رفت تا مدل KNN بتواند ناهمگنی فضایی را بهتر درک کند و دقت پیش‌بینی را افزایش دهد. شاخص‌های کلیدی پوشش گیاهی از تصاویر سری زمانی سنتینل ۲ استخراج و ویژگی‌های مهندسی‌شده برای غنی‌سازی داده‌ها ایجاد شدند. مدل پیشنهادی در مجموعه آزمون به ضریب تبیین (R^2) برابر با ۰/۸۷۰۶ و جذر میانگین مربعات خطا (RMSE) معادل ۷/۸۰ تن در هکتار دست یافت. تحلیل اهمیت ویژگی‌ها نشان داد که متغیرهای مهندسی‌شده، به‌ویژه بهره‌وری آب، از پیش‌بینی‌کننده‌های اصلی عملکرد هستند. نتایج این پژوهش نشان می‌دهد که تلفیق داده‌های ماهواره‌ای با مدل KNN ساده اما اثربخش، ابزاری شفاف و عملی برای پشتیبانی تصمیم‌گیری در کشاورزی دقیق فراهم می‌آورد.
پیش‌بینی عملکرد، خوشه‌بندی K-Means، رگرسیون KNN، سنتینل ۲، نیشکر.	

استناد: جادری، فریال، منجزی، نسیم، (۱۴۰۴) پیش‌بینی عملکرد نیشکر با استفاده از شاخص‌های پوشش‌های گیاهی ماهواره‌ای سنتینل ۲ و الگوریتم‌های خوشه‌بندی

K-MEANS و رگرسیون KNN، مجله مهندسی بیوسیستم ایران / ایران، ۵۶ (۴)، ۹۱-۱۰۵.

<https://doi.org/10.22059/ijbse.2025.403177.665620>

© نویسندگان.

ناشر: مؤسسه انتشارات دانشگاه تهران.

DOI: <https://doi.org/10.22059/ijbse.2025.403177.665620>

مقدمه

نیشکر (*Saccharum officinarum*) یکی از محصولات استراتژیک جهان به‌شمار می‌رود که نه تنها منبع اصلی تولید شکر است، بلکه به‌عنوان ماده‌ی اولیه‌ی کلیدی در تولید سوخت‌های زیستی نیز نقش دارد. مدیریت کارآمد و علمی مزارع نیشکر، نقشی اساسی در پایداری اقتصادی و حفظ امنیت غذایی جهانی ایفا می‌کند. با این حال، کشاورزان و صنایع مرتبط همواره با چالش‌هایی همچون بهینه‌سازی مصرف نهاده‌ها (به‌ویژه آب و کود) و مقابله با پیامدهای تغییرات اقلیمی روبه‌رو هستند. از این رو، پیش‌بینی دقیق و به‌موقع عملکرد پیش از برداشت برای برنامه‌ریزی راهبردی، مدیریت زنجیره‌ی تأمین و افزایش بهره‌وری مزرعه امری ضروری است. روش‌های سنتی برآورد عملکرد معمولاً نیازمند زمان و نیروی کار زیاد بوده و تنها در محدوده‌های کوچک قابل اجرا هستند؛ از این رو نمی‌توانند پاسخ‌گوی نیازهای عملیاتی در مقیاس بزرگ باشند. این محدودیت‌ها ضرورت استفاده از رویکردهای نوین، داده‌محور و مقیاس‌پذیر را برجسته می‌سازد. در سال‌های اخیر، فناوری سنجش از دور با فراهم‌سازی داده‌های چندزمانه و با قدرت تفکیک بالا، تحول چشمگیری در پایش محصولات کشاورزی ایجاد کرده است. تصاویر ماهواره‌ای سری سنتینل ۲ به‌ویژه به دلیل توانایی ثبت داده‌های چندطیفی، امکان ارزیابی مستمر وضعیت سلامت و رشد گیاه را از طریق شاخص‌های پوشش گیاهی فراهم می‌کنند. Li et al. (2023) رابطه‌ی معناداری میان شاخص‌های مشتق‌شده از داده‌های ماهواره‌ای و عملکرد واقعی محصول گزارش کرده‌اند.

رویکردهای مدل‌سازی این روابط از مدل‌های آماری ساده تا روش‌های پیشرفته‌ی یادگیری ماشین گسترش یافته‌اند (de França e Silva et al., 2024). اگرچه الگوریتم‌هایی نظیر جنگل تصادفی، ماشین بردار پشتیبان و شبکه‌های عصبی عمیق عملکرد قابل توجهی از خود نشان داده‌اند، اما بیشتر پژوهش‌ها تنها بر داده‌های طیفی تکیه کرده‌اند و تأثیر ویژگی‌های زراعی مهندسی‌شده، که بیانگر کارایی مدیریتی مزارع هستند، نادیده گرفته شده است. بنابراین، ترکیب هوشمندانه‌ی شاخص‌های پوشش گیاهی و ویژگی‌های مدیریتی می‌تواند گامی مؤثر در جهت بهبود دقت و تفسیرپذیری مدل‌های پیش‌بینی عملکرد نیشکر فراهم آورد.

هدف این پژوهش توسعه یک مدل رگرسیونی K-نزدیک‌ترین همسایه (KNN) مقاوم و قابل تفسیر برای پیش‌بینی عملکرد نیشکر پیش از برداشت با تلفیق تصاویر ماهواره‌ای سنتینل ۲ و ویژگی‌های زراعی مهندسی‌شده است. در این مدل، شاخص‌های کلیدی مانند کارایی مصرف آب و کود به‌عنوان متغیرهای پیش‌بینی‌کننده مهم در نظر گرفته شده‌اند و خوشه‌بندی K-میانگین (K-means) برای طبقه‌بندی مزارع به گروه‌های همگن به کار گرفته شده است تا ناهمگنی فضایی بهتر شناسایی شده و دقت پیش‌بینی افزایش یابد. این رویکرد ترکیبی، امکان ارائه پیش‌بینی‌های دقیق و قابل تفسیر عملکرد نیشکر را فراهم می‌کند، اهمیت داده‌های مدیریتی مزارع را برجسته می‌سازد و با حفظ شفافیت، کارایی محاسباتی و قابلیت اجرا در مقیاس عملیاتی، عملکردی رقابتی در مقایسه با روش‌های پیچیده‌تر ارائه می‌کند.

پیشینه تحقیق

به‌کارگیری فناوری‌های سنجش از دور و یادگیری ماشین به ستون فقرات کشاورزی دقیق مدرن تبدیل شده است و نقشی اساسی در پیش‌بینی عملکرد محصولات استراتژیک، به‌ویژه نیشکر، ایفا می‌کند. در این حوزه، مطالعات علمی رویکردهای متنوعی را معرفی کرده‌اند که می‌توان آن‌ها را بر اساس منبع داده (مانند تصاویر ماهواره‌ای یا داده‌های پهبادی)، سطح پیچیدگی مدل‌های تحلیلی و پارامترهای کشاورزی مورد بررسی، به‌صورت نظام‌مند طبقه‌بندی کرد. این بخش مرور انتقادی این رویکردها را ارائه می‌کند تا جایگاه و نوآوری پژوهش حاضر به‌روشنی تبیین شود.

مطالعات قبلی رویکردهای متنوعی را معرفی کرده‌اند که بر اساس منبع داده (ماهواره یا پهباد)، سطح پیچیدگی مدل‌های تحلیلی و پارامترهای کشاورزی مورد بررسی، قابل دسته‌بندی هستند. با این حال، بسیاری از این پژوهش‌ها محدود به شاخص‌های استاندارد گیاهی نظیر NDVI و EVI بوده و در پوشش‌های گیاهی متراکم دچار اثر اشباع می‌شوند، همچنین حساسیت کافی به تغییرات مدیریتی مزرعه نشان نمی‌دهند.

مدل‌های گروهی یادگیری ماشین برای تحلیل داده‌های ماهواره‌ای، مانند مطالعه Taravat et al. (2024) و Iftikhar et al. (2024)،

1. K-Nearest Neighbor (KNN)
2. K-means clustering
3. Normalized difference vegetation index
4. Enhanced Vegetation Index

توانایی مدیریت روابط پیچیده و غیرخطی و کاهش خطر برآزش بیش از حد را نشان داده‌اند، اما اغلب این مدل‌ها به چرخه فنولوژی محصول و زمان‌بندی مراحل رشد توجه کافی نداشته‌اند، در حالی که پژوهش (Dimov et al. (2022 نشان دادند که اطلاعات سری‌های زمانی شاخص‌های پوشش گیاهی در طول فصل رشد ارتباط مستقیم با عملکرد نهایی دارند. این موضوع نشان می‌دهد که ادغام بعد زمانی برای افزایش دقت پیش‌بینی ضروری است. در بعد روش‌شناختی، مطالعات اخیر مانند (Vasconcelos et al. (2025 و (Akbarian et al. (2022 نشان داده‌اند که تمرکز بر مرحله رشد اولیه و استفاده از اعتبارسنجی متقابل چندساله می‌تواند پایداری مدل را افزایش دهد. علاوه بر این، کانال (Canal Filho et al. (2025 یک مدل یادگیری ماشین تطبیقی چندساله توسعه دادند که توانایی تعدیل روندهای بلندمدت و ارتقای دقت پیش‌بینی را دارد. با این حال، بسیاری از مدل‌های پیچیده گروهی یا یادگیری عمیق نیازمند خطوط پردازش داده و منابع محاسباتی سنگین هستند که کاربرد عملی را محدود می‌کند. ادغام داده‌های چندمنظوره نیز پتانسیل بالایی دارد؛ به‌عنوان مثال، (Zhu et al. (2023 با ترکیب داده‌های نوری و راداری سنتینل ۲ به برآوردهای دقیق عملکرد دست یافتند و (Alemán-Montes et al. (2023 نشان دادند که وضوح مکانی مختلف می‌تواند درک بهتری از ناهمگنی‌های مزرعه‌ای ارائه کند. با این حال، چنین رویکردهایی اغلب پیچیده و نیازمند منابع محاسباتی قابل توجه هستند. پلتفرم‌های جایگزین مانند پهپادها امکان تحلیل‌های دقیق مزرعه‌ای فراهم می‌کنند، اما محدودیت مقیاس و هزینه بالای عملیاتی دارند. اگرچه پلتفرم‌های ماهواره‌ای برای پایش منطقه‌ای در مقیاس بزرگ بی‌رقیب هستند، سایر پلتفرم‌ها نیز مزایای متمایزی ارائه می‌دهند. به‌عنوان نمونه، هواپیماهای بدون سرنشین (پهپادها) وضوح فضایی بسیار بالا در مقیاس سانتی‌متر فراهم می‌کنند و امکان تحلیل دقیق در سطح مزرعه را ایجاد می‌نمایند. مطالعه (Tanut et al. (2021 از این قابلیت برای توسعه مدلی با دقت بالا در پیش‌بینی عملکرد مبتنی بر تصاویر پهپادی استفاده می‌کند. با این حال، روش‌های مبتنی بر پهپاد معمولاً به محدوده‌های کوچک‌تر محدود می‌شوند و هزینه‌های عملیاتی بالاتر و نیاز به پردازش داده‌های سنگین‌تری را به همراه دارند. این یافته‌ها ضرورت بهره‌گیری از داده‌های ماهواره‌ای را برای پایش مقیاس‌پذیر، مقرون‌به‌صرفه و کارآمد برجسته می‌سازد.

کاربرد سنجش از دور و یادگیری ماشین فراتر از پیش‌بینی عملکرد، در سایر پارامترهای کشاورزی نیز به‌خوبی نشان داده شده است. به‌عنوان مثال، (Suwanlee et al. (2024 با استفاده از ترکیب داده‌های سنتینل و الگوریتم‌های یادگیری ماشین، زیست‌توده روی زمین و ذخایر کربن مزارع نیشکر را با موفقیت برآورد کرد. این یافته نشان می‌دهد که ارتباط بنیادی میان داده‌های طیفی و ویژگی‌های بیوفیزیکی گیاه، قابلیت بهره‌برداری برای طیف وسیعی از ارزیابی‌های زیست‌محیطی و کشاورزی را دارد. افزون بر این، اصول بنیادین این رویکردها در گونه‌های مختلف محصول قابل انتقال است. غفاریان‌المیری و همکاران (۱۴۰۳) نیز نشان دادند که به‌کارگیری تکنیک‌های سنجش از دور مبتنی بر ماهواره برای برآورد عملکرد شالیزارهای برنج محصولی با ساختار تاج پوشش و چرخه رشد کاملاً متفاوت از نیشکر نتایج موثری به همراه دارد. موفقیت این مطالعات، بر استحکام و انعطاف‌پذیری رویکردهای مبتنی بر یادگیری ماشین تأکید می‌کند. بنابراین، پژوهش حاضر با شناسایی این محدودیت‌ها و خلأها، با تمرکز بر استخراج ویژگی‌های مهندسی‌شده از داده‌های سنتینل ۲ و طبقه‌بندی مکانی مبتنی بر خوشه‌بندی K-means، به دنبال ارائه مدلی دقیق، مقیاس‌پذیر و قابل تفسیر برای پیش‌بینی عملکرد نیشکر است، که هم حساسیت لازم به مدیریت مزرعه و ناهمگنی‌های مکانی و زمانی را دارد و هم کاربرد عملی آن امکان‌پذیر است.

مواد و روش‌ها

این پژوهش شامل چهار مرحله اولیه است: (۱) جمع‌آوری و پیش‌پردازش داده‌ها، (۲) مهندسی جامع ویژگی، (۳) خوشه‌بندی نمودارهای مزرعه نظارت نشده برای مدیریت ناهمگنی فضایی، و (۴) پیش‌بینی عملکرد نظارت شده با استفاده از مدل K-نزدیکترین همسایه، که با یک روش گروهی اعتبارسنجی می‌شود. هر مرحله در زیربخش‌های بعدی شرح داده شده است (شکل ۱).

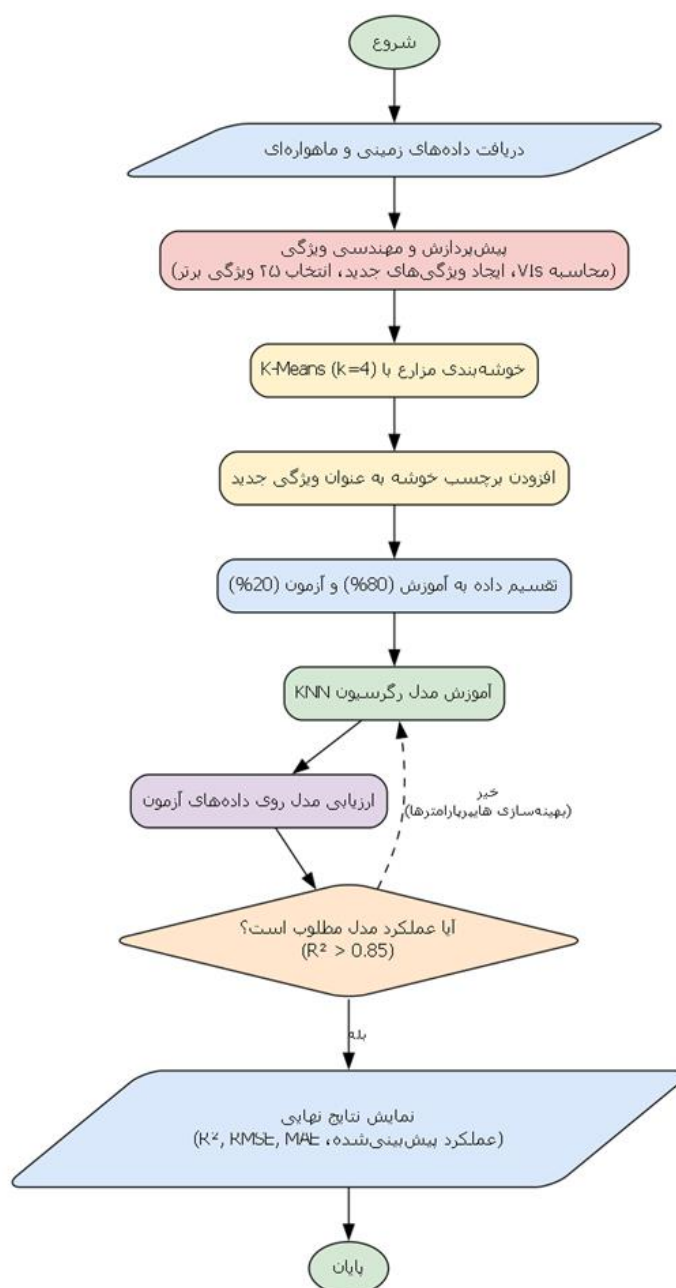
موقعیت محل اجرای تحقیق

این پژوهش در شرکت کشت و صنعت نیشکر دهخدا در سال‌های زراعی ۱۳۹۶-۱۴۰۳ اجرا گردید. کشت و صنعت نیشکر دهخدا در شمال شهرستان اهواز، حد فاصل جاده اندیمشک تا رودخانه کارون، ۲۳ کیلومتر ۲۳ از جاده معروف به دغاغله واقع گردیده است که در موقعیتی با عرض جغرافیایی ۳۱ درجه و ۲۵ دقیقه شمالی و طول جغرافیایی ۴۸ درجه و ۴۰ دقیقه شرقی قرار دارد. خالص اراضی این کشت و صنعت ۱۱۳۶۹ هکتار و ظرفیت اسمی تولید در آن به ترتیب ۱۰۰ هزارتن شکر خام، ۱۷۵ هزارتن شکر سفید و ۳۰ الی ۳۸ هزارتن ملاس و

فرآورده‌های آن و ۳۳۰ هزار تن باگاس است.

شرح مجموعه داده‌های مورد استفاده

داده‌های زراعی و مدیریتی: مجموعه داده‌های زمینی شامل سوابق ۲۴۱۷ قطعه مزرعه است که دوره مورد مطالعه سال‌های کشت ۱۳۹۶ تا ۱۴۰۳ را پوشش می‌دهد. متغیرهای کلیدی شامل بازده ناخالص نهایی (تن در هکتار)، مصرف آب (مترمکعب در هکتار)، مصرف کل کود (کیلوگرم در هکتار) و هدایت الکتریکی خاک^۱ (EC) هستند. پیش از تحلیل، تمامی داده‌های مربوط به قطعات مزرعه از نظر هویتی غیرقابل شناسایی شده تا حفظ حریم خصوصی و محرمانه بودن اطلاعات تضمین شود.



شکل ۱- چارچوب پیشنهادی برای پیش‌بینی عملکرد نیشکر

تصاویر ماهواره‌ای: در این پژوهش از تصاویر چندزمانه سطح 2A (تصحیح جوی) از ابزار چندطیفی ماهواره‌ای^۲ (MSI) سنتینل ۲

استفاده شد. بر اساس یافته‌های Vasconcelos et al. (2025) که بیشترین همبستگی داده‌های مرحله رشد اولیه نیشکر با عملکرد نهایی را نشان می‌دهند، تصاویر برای هر سال در طول فصل اوج رشد (شهریور تا آذر) جمع‌آوری شدند.

پیش‌پردازش داده‌ها و مهندسی ویژگی

قبل از آموزش مدل، مجموعه داده یکپارچه چندین مرحله پیش‌پردازش را پشت سر گذاشت. مقادیر گمشده در ستون‌های عددی با استفاده از میانگین ستون مربوطه محاسبه شد. سپس همه ویژگی‌ها با استفاده از StandardScaler استاندارد شدند. برای افزایش قدرت پیش‌بینی مدل، مهندسی ویژگی گسترده‌ای انجام شد.

شاخص‌های پوشش گیاهی^۱ (VIs): برای ارزیابی کمی سلامت و رشد محصول نیشکر، سه شاخص کلیدی پوشش گیاهی از باندهای طیفی ماهواره سنتینل ۲ محاسبه شدند. این شاخص‌ها به دلیل حساسیت بالا به ویژگی‌های بیوفیزیکی گیاه انتخاب شدند.

شاخص تفاضلی نرمال‌شده پوشش گیاهی^۲ (NDVI): این شاخص، پرکاربردترین معیار برای سنجش میزان سرسبزی و تراکم پوشش گیاهی است. مقادیر بالاتر NDVI نشان‌دهنده پوشش گیاهی سالم‌تر است. این شاخص با استفاده از بازتاب در باندهای فرورسرخ نزدیک (NIR) و قرمز (Red) طبق رابطه (۱) محاسبه می‌شود.

$$NDVI = \frac{(NIR - Red)}{(NIR + Red)} \quad \text{رابطه (۱)}$$

کاربرد گسترده این شاخص برای پایش محصولات کشاورزی در مطالعات متعددی تأیید شده است (Xue & Su, 2017).

شاخص پوشش گیاهی سبز نرمال‌شده^۳ (GNDVI): این شاخص از باند سبز (Green) به جای قرمز استفاده می‌کند و نسبت به NDVI به غلظت کلروفیل در برگ‌ها حساس‌تر است. این ویژگی آن را برای تشخیص تنش‌های غذایی گیاه مناسب می‌سازد که با استفاده از رابطه (۲) محاسبه می‌شود.

$$GNDVI = \frac{(NIR - Green)}{(NIR + Green)} \quad \text{رابطه (۲)}$$

این شاخص به دلیل حساسیت بالا به کلروفیل، به ویژه در مراحل میانی رشد محصول بسیار مفید است (Haboudane et al., 2002).

ویژگی‌های مهندسی شده: پنج ویژگی جدید برای به دست آوردن کارایی و برهمکنش‌های زراعی ایجاد شد: بهره‌وری آب^۴، کارایی کود^۵، نسبت آبیاری به کود^۶، شاخص سلامت پوشش گیاهی^۷ و برهمکنش خاک و پوشش گیاهی^۸ که به ترتیب با استفاده از روابط (۳ تا ۷) محاسبه می‌شود.

$$WUE = \frac{Y}{W} \quad \text{رابطه (۳)}$$

Y = بازده ناخالص نهایی بر حسب تن / هکتار (t/ha)

W = مصرف آب بر حسب متر مکعب / هکتار (m³/ha)

$$FUE = \frac{Y}{F} \quad \text{رابطه (۴)}$$

F = کود کل بر حسب کیلوگرم / هکتار (kg/ha)

$$IFR = \frac{W}{F + \epsilon} \quad \text{رابطه (۵)}$$

حالت صفر یا حالت گمشده ϵ

$$VHI = \frac{NDVI_i - NDVI_{min}}{NDVI_{max} - NDVI_{min}} \quad \text{رابطه (۶)}$$

1. Vegetation Indices

2. Normalized Difference Vegetation Index

3. Green Normalized Difference Vegetation Index

4. Water Use Efficiency

5. Fertilizer Use Efficiency

6. Irrigation-to-Fertilizer Ratio

7. Vegetation Health Index

8. Soil-Vegetation Interaction Index

$NDVI_i$ = مقدار شاخص در مقطع زمانی مورد نظر

$NDVI_{min}$ = حداقل مقدار ثبت شده در طول دوره‌ی مورد بررسی برای آن منطقه

$NDVI_{max}$ = حداکثر مقدار ثبت شده در طول دوره‌ی مورد بررسی برای آن منطقه

رابطه ۷)

$$SVII = \frac{NDVI}{GNDVI}$$

انتخاب ویژگی: الگوریتم SelectKBest با استفاده از تابع امتیازدهی رگرسیون F، برای انتخاب ۲۵ ویژگی برتر آموزنده برای

آموزش مدل استفاده شد.

چارچوب پیش‌بینی خوشه‌بندی و رگرسیون

این چارچوب از یک رویکرد ترکیبی دوماه‌ای برای شناسایی الگوهای فضایی زیربنایی استفاده می‌کند و در ادامه به پیش‌بینی عملکرد می‌پردازد.

خوشه‌بندی مزرعه بدون نظارت با K-means: برای محاسبه ناهمگنی فضایی، ابتدا از الگوریتم K-means برای تقسیم مزارع

به گروه‌های همگن استفاده شد. K-means مجموع مربعات درون خوشه‌ای (اینرسی) را به حداقل می‌رساند که از طریق رابطه (۸) محاسبه می‌شود.

جایی که J تابع هدف است، K تعداد خوشه‌ها است (از طریق روش آرنج تعیین می‌شود)، $x_i(j)$ نقطه داده i در خوشه j و c_j مرکز خوشه j است. برچسب خوشه به دست آمده به عنوان یک ویژگی طبقه‌بندی جدید اضافه شد.

پیش‌بینی عملکرد نظارت شده: مدل مورد استفاده، الگوریتم K-نزدیک‌ترین همسایه است. این یک روش ناپارامتری است که

عملکرد یک مزرعه مشخص را بر اساس میانگین عملکرد K مزرعه‌ای که بیشترین شباهت را به آن دارند، پیش‌بینی می‌کند. برای سنجش این شباهت، از فاصله اقلیدسی در فضای ویژگی‌ها استفاده می‌شود.

رابطه ۸)

$$d(p, q) = \sqrt{\sum_{i=1}^n (p_i - q_i)^2}$$

که در آن $d(p, q)$ فاصله بین قطعات مزرعه p و q است.

معیارهای آموزش و ارزیابی مدل

مجموعه داده‌ها با استفاده از رویکرد طبقه‌بندی شده بر اساس برچسب‌های خوشه‌ای K-means به مجموعه‌های آموزشی (۸۰ درصد) و آزمون (۲۰ درصد) تقسیم شدند. برای اطمینان از استحکام مدل و جلوگیری از برازش بیش از حد، از یک تکنیک اعتبارسنجی متقابل ۵ برابری بر روی داده‌های آموزشی استفاده شد. عملکرد مدل‌ها با استفاده از سه معیار استاندارد کمی شد.

ضریب تبیین (R^2)

رابطه ۹)

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

رابطه ۱۰)

$$J = \sum_{j=1}^k \sum_{i=1}^n \|x_i^{(j)} - c_j\|^2$$

جذر میانگین مربعات خطا ($RMSE$)

رابطه ۱۱)

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

1. Elbow Method
2. Correlation coefficient
3. Root mean square error

میانگین خطای مطلق (MAE)

رابطه ۱۲)

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

جایی که y_i بازده ناخالص نهایی واقعی است، $\hat{y}_i$ بازده پیش‌بینی شده و $\bar{y}$ میانگین بازده واقعی است. این معیارها برای ارائه یک ارزیابی جامع انتخاب شده‌اند. R^2 خوب بودن برازش را ارزیابی می‌کند، جذر میانگین مربعات خطا، خطاهای بزرگ را سنگین تر جریمه می‌کند، و میانگین خطای مطلق معیاری به راحتی قابل تفسیر از مقدار متوسط خطا ارائه می‌دهد.

نتایج و بحث

این بخش، یافته‌های تجربی این مطالعه را تشریح می‌کند که به‌طور ویژه بر عملکرد مدل پیشنهادی K-نزدیکترین همسایه‌ها متمرکز است. در ابتدا، یک دیدگاه کلی آماری از مجموعه داده ارائه می‌شود و سپس نتایج خوشه‌بندی نمودار مزرعه که به عنوان ورودی اصلی مدل عمل می‌کند، تشریح می‌شود. در نهایت، جزئیات مربوط به دقت پیش‌بینی مدل KNN به طور کامل توضیح داده می‌شود.

بررسی اجمالی مجموعه داده

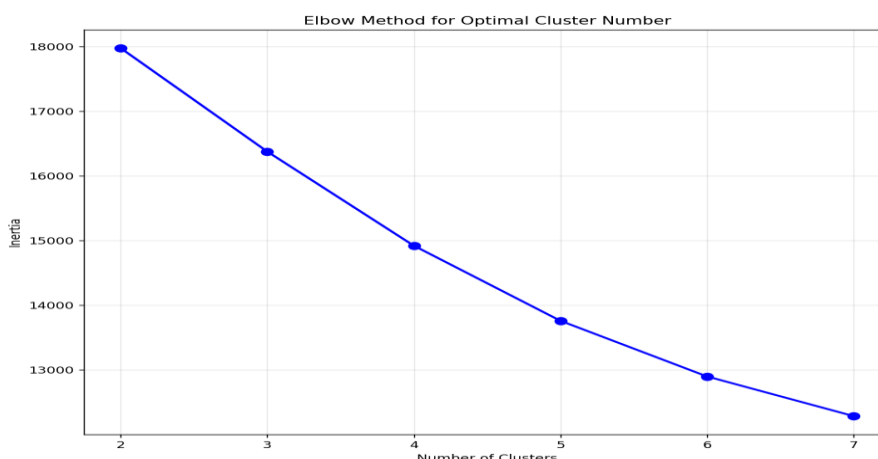
در این مطالعه از یک مجموعه داده ترکیبی شامل ۲۴۱۷ قطعه نمونه مزرعه از سال‌های کشت ۱۳۹۶ تا ۱۴۰۳ استفاده شد. این مجموعه داده دو منبع اصلی را ادغام می‌کند: داده‌های زراعی زمینی (به عنوان مثال، عملکرد، ورودی‌های آب/کود) که مستقیماً از مزارع جمع‌آوری شده است، و داده‌های سنجش از دور به دست آمده از تصاویر ماهواره‌ای سنتینل ۲. جدول (۱) ویژگی‌های آماری متغیرهای کلیدی را خلاصه می‌کند. داده‌ها طیف گسترده‌ای از عملکرد ناخالص نهایی (۱۴/۵۸ تا ۱۴۷/۱۳ تن در هکتار) را نشان می‌دهد که نشان‌دهنده تفاوت‌های بهره‌وری قابل توجهی است که مدل باید پیش‌بینی آن را بیاموزد.

جدول ۱- آمار توصیفی متغیرهای کلیدی

متغیر	میانگین	انحراف معیار	حداقل	حداکثر
بازده ناخالص نهایی (تن در هکتار)	۶۷/۱۶	۲۲/۱۷	۱۴/۵۸	۱۴۷/۱۳
مصرف آب (متر مکعب در هکتار)	۱۴۲۰/۹۵	۱۲۹/۰۷	۱۰۰۴/۷۸	۱۹۰۰/۰۰
کود کل (کیلوگرم در هکتار)	۳۴۸/۸۶	۴۷/۷۰	۱۶۰/۲۷	۵۵۳/۳۳
EC خاک	۵/۴۷	۲/۱۱	۱/۱۰	۱۶/۴۰
میانگین NDVI	۰/۰۶	۰/۰۶	-۰/۰۳	۰/۲۴

نتایج خوشه‌بندی مزرعه

به عنوان یک گام مقدماتی برای طبقه‌بندی پلات‌های مزرعه، از الگوریتم K-means استفاده شد. تعداد بهینه خوشه‌ها $k=4$ تعیین شد که با روش آرنج آتوجیه می‌شود (شکل ۲).



شکل ۲- تعیین تعداد بهینه خوشه‌ها ($k=4$) با استفاده از روش آرنج

میانگین ویژگی‌های این چهار خوشه در جدول (۲) خلاصه شده است. پروفایل‌های زراعی متمایز مشاهده می‌شود:

خوشه صفر: کرت‌های با بهره‌وری متوسط ۶۳/۱۳ تن در هکتار با بیشترین میانگین شوری خاک ۶/۱۰

خوشه ۱: کرت‌های با بیشترین عملکرد ۷۶/۷۸ تن در هکتار، با شوری کم و بیشترین میانگین مصرف کود.

خوشه ۲: کرت‌های کم عملکرد ۵۸/۴۷ تن در هکتار که به طور متوسط کمترین کود را دریافت می‌کردند.

خوشه ۳: کرت‌های بسیار کارآمد با عملکرد بالا ۷۱/۵۴ تن در هکتار با وجود نهاده‌های آب و کود متوسط.

جدول ۲- ویژگی‌های خوشه‌های مزرعه شناسایی شده

خوشه	تعداد نمونه‌ها	میانگین عملکرد (تن در هکتار)	میانگین آب (متر مکعب در هکتار)	میانگین کود (کیلوگرم در هکتار)	میانگین شوری خاک
۰	۷۶۱	۶۳/۱۳	۱۴۳۳/۶۲	۳۵۱/۸۳	۶/۱۰
۱	۷۲۱	۷۶/۷۸	۱۳۹۹/۷۵	۳۶۶/۲۶	۴/۵۹
۲	۶۰۹	۵۸/۴۷	۱۴۳۳/۱۲	۳۳۲/۵۴	۵/۵۰
۳	۳۲۶	۷۱/۵۴	۱۴۱۵/۵۰	۳۳۳/۹۶	۵/۹۲

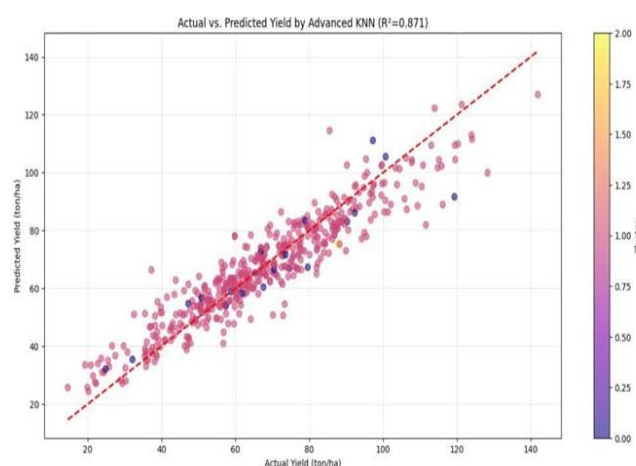
عملکرد پیش‌بینی مدل پیشنهادی KNN

عملکرد مدل KNN در جدول (۳) ارائه شده است. در داده‌های آزمون دیده نشده، مدل به $R^2 = 0.8706$ دست یافت که نشان می‌دهد ۸۷ درصد از واریانس بازده ناخالص نهایی را تبیین می‌کند. جذر میانگین مربعات خطا (RMSE) ۷/۸۰ تن در هکتار و میانگین خطای مطلق (MAE) ۶/۵۸۹۱ تن در هکتار بود.

جدول ۳- ارزیابی عملکرد مدل KNN

معیار ارزیابی	مجموعه آزمون	اعتبارسنجی متقابل ۵ برابر
R^2	۰/۸۷۰۶	۰/۸۱۸۱ ± ۰/۱۹۷
RMSE	۷/۸۰	-
MAE	۶/۵۸۹۱	-

قابلیت پیش‌بینی مدل در شکل (۳) به صورت بصری نمایش داده شده است، به طوری که مقادیر پیش‌بینی شده در برابر مقادیر واقعی ترسیم شده‌اند. توزیع یکنواخت نقاط داده در امتداد خط ۱:۱ نشان‌دهنده دقت بالای مدل KNN است. همچنین، عملکرد مدل نسبت به پژوهش‌های مشابه از جمله رویکرد جنگل تصادفی گزارش شده در مطالعه Taravat et al. (2024) برتری نشان می‌دهد. هرچند مدل KNN ممکن است به بالاترین دقت روش‌های گروهی پیچیده دست نیابد، اما از نظر سادگی، قابلیت تفسیر و سهولت پیاده‌سازی، مزایای چشمگیری دارد.



شکل ۳ - نمودار پراکندگی بازده واقعی نهایی در مقایسه با مقادیر پیش‌بینی شده توسط مدل KNN در مجموعه داده آزمون

بحث

نتایج ارائه شده در بخش پیشین نشان می‌دهد که مدل پیشنهادی K-نزدیک‌ترین همسایه (KNN) ابزاری توانمند برای پیش‌بینی عملکرد نیشکر است و توانسته است به ضریب تبیین بالایی ($R^2 = 0.8706$) بر روی داده‌های آزمون مشاهده‌نشده دست یابد. این بخش به تفسیر یافته‌ها پرداخته، آن‌ها را در چارچوب ادبیات موجود قرار داده و پیامدها و محدودیت‌های پژوهش را بررسی می‌کند. پرسش اصلی تحقیق این بود که آیا مدلی ساده و قابل تفسیر مانند KNN، که با طبقه‌بندی فضایی و ویژگی‌های مهندسی شده تقویت شده است، قادر به پیش‌بینی دقیق عملکرد خواهد بود یا خیر. یافته‌ها این فرضیه را تأیید می‌کنند. مرحله خوشه‌بندی اولیه با بهره‌گیری از الگوریتم K-میانگین (K-means) در تفکیک مزارع به چهار نیمرخ زراعی مجزا (مطابق جدول ۲) مؤثر واقع شد. این تفکیک با فراهم کردن امکان انتخاب همسایگان از گروه‌های همگن‌تر، به بهبود عملکرد مدل کمک کرده است. برای نمونه، مدل توانست به‌خوبی بین مزارع با پتانسیل بالا (خوشه ۱) و مزارع با عملکرد پایین (خوشه ۲) تمایز قائل شود. اگرچه دستیابی مدل KNN به ضریب تبیین کامل در داده‌های آموزشی ($R^2 = 1.0$) نشان‌دهنده تمایل به برازش شدید داده‌ها است، اما عملکرد قوی آن بر روی داده‌های آزمون ($R^2 = 0.8706$) توانایی تعمیم مدل به داده‌های جدید و مشاهده نشده را تأیید می‌کند. در مقایسه با مطالعات پیشین، عملکرد مدل پیشنهادی بسیار رقابتی است. برای نمونه، (Taravat et al., 2024) نتایج قابل توجهی در پیش‌بینی عملکرد نیشکر با بهره‌گیری از مدل جنگل تصادفی پیچیده و داده‌های سنتینل ۲ ارائه کرده است. اگر چه روش‌های گروهی ممکن است دقت اوج بالاتری ارائه دهند، مدل KNN مزایای قابل توجهی از نظر سادگی و تفسیرپذیری دارد. این ویژگی، KNN را به ابزاری کاربردی‌تر برای استفاده در سیستم‌های پشتیبانی تصمیم‌گیری در شرایط واقعی تبدیل می‌کند.

علاوه بر این، استفاده از پلتفرم مبتنی بر ماهواره در این مطالعه مزایای قابل توجهی نسبت به روش‌های مبتنی بر پهپاد ارائه می‌دهد، همانند رویکرد گزارش شده در تحقیق (Tanut et al., 2021). در حالی که پهپادها وضوح فضایی بالاتری ارائه می‌دهند، روش پیشنهادی از نظر مقیاس‌پذیری و مقرون به‌صرفه بودن با مأموریت سنتینل ۲ مزیت دارد، و آن را برای نظارت مداوم و در مقیاس بزرگ بر مناطق کشاورزی و صنعتی بدون پیچیدگی‌های عملیاتی پهپادها فراهم می‌کند.

جدول (۴) مقایسه جامعی از عملکرد مدل پیشنهادی با روش‌های پیشرفته اخیر در پیش‌بینی عملکرد نیشکر ارائه می‌دهد.

جدول ۴- تجزیه و تحلیل مقایسه‌ای روش‌های پیش‌بینی عملکرد نیشکر

مطالعات اخیر	روش	منابع داده	ضریب تبیین	جذر میانگین مربعات خطا (تن در هکتار)	ویژگی‌های کلیدی	تفسیرپذیری
پژوهش حاضر	K-میانگین	داده‌های مزرعه + همسایه (KNN) سنتینل ۲	۰/۸۷۰۶	۷/۸۰	ویژگی‌های مهندسی شده، بهره‌وری آب	بالا
تراوات و همکاران (۲۰۲۴)	جنگل تصادفی (RF)	سنتینل ۲	۰/۸۵۱۲	۹/۲۳	سری‌های زمانی NDVI, EVI	متوسط
افتخار و همکاران (۲۰۲۴)	شبکه عصبی عمیق (DNN)	لندست ۸ + داده‌های هواشناسی	۰/۸۷۳۴	۷/۹۵	شاخص‌های چند زمانه	کم
واسکونسوس و همکاران (۲۰۲۵)	ماشین بردار پشتیبان (SVM)	سنتینل ۲	۰/۸۱۰۲	۱۰/۴۵	شاخص‌های فاز رشد	متوسط
تانوت و همکاران (۲۰۲۱)	شبکه عصبی پیچشی (CNN)	تصاویر پهپاد	۰/۹۱۵۶	۶/۳۲	RGB با وضوح بالا	کم
زو و همکاران (۲۰۲۳)	جنگل تصادفی + همجوشی داده‌ها	سنتینل ۲و۱	۰/۸۸۲۳	۸/۱۵	نوری + SAR	متوسط
کانال فلیو و همکاران (۲۰۲۵)	یادگیری ماشین تطبیقی (AML)	سری زمانی تصاویر MODIS	۰/۸۴۴۵	۹/۶۷	ویژگی‌های زمانی	متوسط

همان‌طور که در جدول (۴) نشان داده شده است، اگرچه رویکرد مبتنی بر KNN در مقایسه با روش‌های یادگیری عمیق (Tanut et al., 2021) یا تکنیک‌های گروهی پیچیده، بالاترین سطح دقت مطلق را به دست نمی‌آورد، اما چندین مزیت متمایز ارائه می‌دهد. پیش‌بینی‌های KNN قابلیت تفسیر بالایی دارند، زیرا می‌توان آن‌ها را مستقیماً به نمونه‌های تاریخی مشابه ردیابی کرد و فرآیند تصمیم‌گیری را برای ذینفعان شفاف می‌سازد. علاوه بر این، این روش به منابع محاسباتی کمتری نیاز دارد و در مقایسه با رویکردهای یادگیری عمیق، برای استقرار عملیاتی در محیط‌های با منابع محدود مناسب است. با وجود سادگی، مدل پیشنهادی عملکردی رقابتی با روش‌های پیچیده‌تر ارائه می‌کند و در عین حال اثربخشی رویکرد مهندسی ویژگی‌ها را تأیید می‌نماید.

نتایج این پژوهش به‌خوبی به پرسش اصلی تحقیق پاسخ می‌دهد و نشان می‌دهد که یک مدل ساده و مبتنی بر نمونه، که با طبقه‌بندی فضایی و ویژگی‌های مهندسی شده متناسب با حوزه مورد مطالعه تقویت شده است، می‌تواند به‌عنوان ابزاری قدرتمند در کشاورزی دقیق مورد استفاده قرار گیرد. همچنین، عملکرد قابل توجه راندمان آب به‌عنوان متغیر پیش‌بینی، مطابق با یافته‌های این مطالعه، با نتایج گزارش شده توسط Li et al. (2023) و de França e Silva et al. (2024) هم‌راستا است که بر نقش کلیدی مدیریت آبیاری در بهینه‌سازی عملکرد محصول تأکید کرده‌اند.

برای تحقیقات آینده، چندین مسیر امیدوارکننده وجود دارد. ادغام داده‌های چندوجهی می‌تواند محدودیت‌های فعلی مطالعه را برطرف کند؛ الهام از موفقیت Zhu et al. (2023) در همجوشی داده‌های سنتینل ۱ و ۲، همراه با ترکیب داده‌های هواشناسی پویا از منابعی مانند تحلیل مجدد ERA5 و ویژگی‌های اضافی خاک (pH و محتوای ماده آلی)، می‌تواند دقت مدل را افزایش دهد، و پژوهش (Alemán-Montes et al., 2023) نشان داده است که چنین ادغام چندوجهی می‌تواند R^2 را تا ۱۵٪ بهبود دهد. استفاده از چارچوب‌های یادگیری چندوظیفه‌ای با الهام از Sun et al. (2024) و پیشرفت‌های اخیر در هوش مصنوعی کشاورزی، مسیر دیگری برای توسعه مدل فراهم می‌آورد؛ چنین رویکردی قادر است هم‌زمان پارامترهای متعددی از جمله عملکرد محصول، محتوای قند و تجمع زیست‌توده را پیش‌بینی کند و ارزیابی جامع‌تری از ارزش محصول ارائه دهد. ترکیب یادگیری عمیق زمانی با حفظ مزیت تفسیرپذیری مدل KNN نیز می‌تواند با کاوش رویکردهای هیبریدی، که KNN را با معماری‌های شبکه‌های حافظه کوتاه‌مدت بلند و شبکه‌های کانولوشن زمانی ادغام می‌کنند، ارتقا یابد؛ این رویکردها قابلیت‌های پیشرفته‌ای برای مدل‌سازی پویایی رشد محصول فراهم می‌کنند و مطالعه اخیر غفاریان مالمیری و همکاران (۱۴۰۳) نشان داده است که این روش‌های ترکیبی می‌توانند تفسیرپذیری مدل را حفظ کرده و دقت پیش‌بینی را ۸ تا ۱۲ درصد افزایش دهند. همچنین، یادگیری انتقال و انطباق دامنه مسیر دیگری برای تحقیقات آینده است؛ با توجه به موفقیت روش‌شناسی حاضر، بررسی قابلیت انتقال مدل به سایر مناطق کشت نیشکر با شرایط آب‌وهوایی و شیوه‌های کشت متفاوت اهمیت دارد و تکنیک‌های انطباق دامنه، همان‌طور که Akbarian et al. (2022) نشان داده‌اند، می‌توانند مدل را قادر سازند تا در زمینه‌های جغرافیایی مختلف تعمیم یابد. در نهایت، توسعه یک سیستم عملیاتی بلادرنگ مبتنی بر ابر که مدل حاضر را با جریان‌های داده سنتینل ۲ تقریباً بلادرنگ ادغام کند، می‌تواند این تحقیق را به ابزاری عملی برای کشاورزان و تصمیم‌گیرندگان تبدیل کند و امکان ارائه پیش‌بینی عملکرد هفتگی در طول فصل رشد و بهینه‌سازی پویا منابع را فراهم آورد.

نتیجه گیری

در این مطالعه، چارچوبی مبتنی بر K-نزدیک‌ترین همسایه (KNN) برای پیش‌بینی عملکرد نیشکر با بهره‌گیری از ترکیبی از تصاویر ماهواره‌ای سنتینل ۲ و داده‌های مدیریت مزرعه ارائه و اعتبارسنجی شد. با انجام خوشه‌بندی K-means بر روی مزارع و سپس استفاده از الگوریتم KNN، مدل توانست روابط پیچیده بین ویژگی‌های ورودی و خروجی محصول نهایی را به‌طور موفقیت‌آمیز استخراج کند. مدل پیشنهادی بر روی داده‌های آزمون مشاهده‌نشده با ضریب تبیین $R^2=0.8706$ عملکرد پیش‌بینی قوی از خود نشان داد، که مؤید این است که این رویکرد یک راه‌حل مؤثر، مقیاس‌پذیر و با قابلیت تفسیر بالا برای برآورد عملکرد قبل از برداشت ارائه می‌دهد. در نتیجه، این مطالعه نشان می‌دهد که ترکیب یادگیری ماشین سنتی با مهندسی ویژگی‌های هوشمندانه و تحلیل فضایی می‌تواند نتایجی رقابتی با رویکردهای پیچیده‌تر ارائه دهد و هم‌زمان شفافیت و تفسیرپذیری لازم برای کاربردهای عملی در کشاورزی را حفظ کند. یافته‌های این مطالعه پیامدهای عملی قابل توجهی برای کشاورزی دقیق دارد. چارچوب توسعه‌یافته می‌تواند به‌عنوان یک ابزار پشتیبان تصمیم‌گیری برای مدیران مزرعه

به کار گرفته شود تا مناطق ضعیف در اوایل فصل شناسایی شوند، تخصیص فضایی منابعی مانند آب و کود بهینه شود و پیش‌بینی‌های قابل اعتماد برای زنجیره تأمین و برنامه‌ریزی لجستیکی فراهم گردد. از نظر تئوری، این تحقیق اثربخشی یک رویکرد ترکیبی را نشان می‌دهد که طبقه‌بندی فضایی بدون نظارت را با یک مدل رگرسیون ساده و مبتنی بر نمونه برای یک وظیفه پیش‌بینی زراعی پیچیده هم‌آهنگ می‌کند. با این حال، مطالعه با محدودیت‌هایی همراه است. نخست، مدل بدون در نظر گرفتن داده‌های دینامیک هواشناسی (مانند دما و بارش) یا ویژگی‌های خاک حیاتی مانند pH توسعه یافته است که این عوامل تأثیر زیادی بر رشد محصول دارند. دوم، به عنوان یک مدل ناپارامتریک، عملکرد KNN به کیفیت و نمایندگی داده‌های آموزشی بستگی دارد.

منابع

عربی علی آباد. (۱۴۰۳). شناسایی و برآورد عملکرد مزارع برنج با استفاده از تصاویر غفاریان مالمیری، صابری، محمدعارف، مظفری، غلامعلی، مجله جغرافیا و توسعه، ۲۲(۷۴)، ۲۱۸-۱۷۸. ماهواره‌های و تکنیک‌های سنجش‌ازدور (مطالعه موردی: استان کندز، افغانستان)
<http://dx.doi.org/10.22111/GDIJ.2024.8177>

REFERENCES

- Alemán-Montes, B., Zabala, A., Henríquez, C., & Serra, P. (2023). Modelling Two Sugarcane Agro-Industrial Yields Using Sentinel/Landsat Time-Series Data and Their Spatial Validation at Different Scales in Costa Rica. *Remote sensing*, 15(23), 5476. DOI: <https://doi.org/10.3390/rs15235476>
- Akbarian, S., Xu, C., Wang, W., Ginns, S., & Lim, S. (2022). Sugarcane yields prediction at the row level using a novel cross-validation approach to multi-year multispectral images. *Computers and Electronics in Agriculture*, 198, 107024. DOI: <https://doi.org/10.1016/j.compag.2022.107024>
- Canal Filho, R., Molin, J., Silva, E., Wei, M., & Sterle, L. (2025). Adaptive multi-year machine learning model to predict sugarcane yield. In *Precision agriculture'25* (pp. 716–722). Wageningen Academic. DOI: https://doi.org/10.1163/9789004725232_094
- de França e Silva, N. R., Chaves, M. E. D., Luciano, A. C. d. S., Sanches, I. D. A., de Almeida, C. M., & Adami, M. (2024). Sugarcane yield estimation using satellite remote sensing data in empirical or mechanistic modeling: A systematic review. *Remote sensing*, 16(5), 863. DOI: <https://doi.org/10.3390/rs16050863>
- Dimov, D., Uhl, J. H., Löw, F., & Seboka, G. N. (2022). Sugarcane yield estimation through remote sensing time series and phenology metrics. *Smart Agricultural Technology*, 2, 100046. DOI: <https://doi.org/10.1016/j.atech.2022.100046>
- Ghafarian Malmiri, H. R., Saberi, M. A., Mozafari, Gh. A., & Arabi Aliabad, F. (2024). Rice Paddies Mapping and Yield Estimating using Satellite Images and Remote Sensing Techniques (Case study: Kunduz province, Afghanistan) *Geography and Development*, 22 (74), 187-218. DOI: <http://dx.doi.org/10.22111/GDIJ.2024.8177> (in Persian)
- Haboudane, D., Miller, J. R., Tremblay, N., Zarco-Tejada, P. J., & Dextraze, L. (2002). Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture. *Remote sensing of environment*, 81(2-3), 416-426. DOI: [https://doi.org/10.1016/S0034-4257\(02\)00018-4](https://doi.org/10.1016/S0034-4257(02)00018-4)
- Iftikhar, M., Qadri, S., Nadeem, M., & Nawaz, S. A. (2024). Remote Sensing Based Sugarcane Yield Prediction Model using Artificial Intelligence. *Journal of Computing & Biomedical Informatics*, 6(02), 451–462. DOI: <https://doi.org/10.56979/602/2024>
- Li, H., Di, L., Zhang, C., Lin, L., Guo, L., Zhao, H., ... & Hong, R. (2023), July. A review of remote sensing in sugarcane mapping. In *2023 11th International Conference on Agro-Geoinformatics (Agro-Geoinformatics)* (pp. 1-5). IEEE. DOI: <https://doi.org/10.1109/Agro-Geoinformatics59224.2023.10233506>
- Suwanlee, S. R., Pinasu, D., Som-ard, J., Borgogno-Mondino, E., & Sarvia, F. (2024). Estimating sugarcane aboveground biomass and carbon stock using the combined time series of sentinel data with machine learning algorithms. *Remote sensing*, 16(5), 750. DOI: <https://doi.org/10.3390/rs16050750>
- Sun, J., Sun, C., Li, Z., Qian, Y., & Li, T. (2024). Prediction method of sugarcane important phenotype data based on multi-model and multi-task. *PloS one*, 19(12), e0312444. DOI: <https://doi.org/10.1371/journal.pone.0312444>
- Taravat, A., Abebe, G., Gessesse, B., & Tadesse, T. (2024). Estimation of Sugarcane Yield Using Multi-Temporal Sentinel 2 Satellite Imagery and Random Forest Regression. DOI:

<https://doi.org/10.5194/isprs-archives-XLVIII-4-W9-2024-357-2024>

- Tanut, B., Waranusast, R., & Riyamongkol, P. (2021). High accuracy pre-harvest sugarcane yield forecasting model utilizing drone image analysis, data mining, and reverse design method. *Agriculture*, 11(7), 682. DOI: <https://doi.org/10.3390/agriculture11070682>
- Vasconcelos, J. C. S., Arantes, C. S., Speranza, E. A., Antunes, J. F. G., Barbosa, L. A. F., & Cançado, G. M. d. A. (2025). Predicting Sugarcane Yield Through Temporal Analysis of Satellite Imagery During the Growth Phase. *Agronomy*, 15(4), 793. DOI: <https://doi.org/10.3390/agronomy15040793>
- Xue, J., & Su, B. (2017). Significant remote sensing vegetation indices: A review of developments and applications. *Journal of sensors*, 2017(1), 1353691. DOI: <https://doi.org/10.1155/2017/1353691>
- Zhu, L., Liu, X., Wang, Z., & Tian, L. (2023). High-precision sugarcane yield prediction by integrating 10-m Sentinel-1 VOD and Sentinel-2 GRVI indexes. *European Journal of Agronomy*, 149, 126889. DOI: <https://doi.org/10.1016/j.eja.2023.126889>