

ارزیابی عملکرد مدل‌های مبتنی بر یادگیری ماشین در برآورد ابعاد الگوی توزیع رطوبت تحت آبیاری قطره‌ای

چکیده

امروزه مدل‌های یادگیری ماشین به عنوان جایگزین مناسب مدل‌های تحلیلی، عددی و تجربی جهت تخمین ابعاد الگوی توزیع رطوبت در خاک بدون محدودیت‌هایی مانند تعریف شرایط مرزی یا بازکالیبراسیون مورد توجه قرار گرفته‌اند. از این رو در تحقیق حاضر با استفاده از اطلاعات هشت متغیر ورودی (درصد شن، درصد سیلت، درصد رس، شوری آب، زمان تجمع، حجم آب تجمع، وزن مخصوص ظاهری و هدایت هیدرولیکی اشباع) به ارزیابی عملکرد پنج مدل یادگیری ماشین (CatBoost, XGBoost, RF, SVR, Elnet) برای تخمین عمق نفوذ و عرض سطحی پیاز رطوبتی در دو بافت مختلف خاک (Loamy Sand, Sandy Clay) تحت سیستم آبیاری قطره‌ای سطحی پرداخته شد. در مدل‌های مذکور از روش اعتبارسنجی متقاطع استاندارد با اختلاط تصادفی استفاده شد و عملکرد مدل‌ها با استفاده از شاخص‌های آماری MAE, RMSE, R^2 ارزیابی گردید. نتایج بدست آمده نشان داد که دقیق‌ترین تخمین عمق نفوذ در بافت Sandy Clay (MAE = 0.10 cm, RMSE = 0.27 cm, $R^2 = 0.94$) و عرض سطحی (MAE = 0.91 cm, RMSE = 1.58 cm, $R^2 = 0.98$) در بافت Sandy Clay و در مدل CatBoost مشاهده شد. به‌طوری‌که پایین‌ترین دقت تخمین عمق نفوذ و عرض سطحی با مدل مذکور به ترتیب در بافت‌های Loamy Sand (MAE = 3.22 cm, RMSE = 6.16 cm, $R^2 = 0.16$) و Sandy Clay (MAE = 4.67 cm, RMSE = 6.09 cm, $R^2 = 0.76$) مشاهده گردید. در مجموع نتایج بدست آمده توانایی مدل‌های یادگیری ماشین مبتنی بر روش‌های درخت-پایه را که قادر به یادگیری رفتار پیچیده و غیرخطی بین متغیرها هستند، در برآورد ابعاد پیاز رطوبتی به ویژه در خاک با درصد رس بالا تایید می‌کند.

کلید واژه‌ها: ابعاد الگوی خیس شدگی، الگوریتم CatBoost، شبیه‌سازی، مدل‌های داده محور.

Performance Evaluation of Machine Learning-Based Models in Estimating Soil Moisture Distribution Dimensions under Drip Irrigation

ABSTRACT

In recent years, machine learning models have been increasingly recognized as effective alternatives to analytical, numerical, and empirical models for estimating the dimensions of soil moisture distribution patterns, without requiring constraints such as boundary condition specification or recalibration. Therefore, in the present study, the performance of five machine learning models (CatBoost, XGBoost, RF, SVR, Elnet) was evaluated for estimating infiltration depth and surface width of the wetting bulb in two soil textures (Loamy Sand, Sandy Clay) under surface drip irrigation system.

Eight input variables including sand percentage, silt percentage, clay percentage, water salinity, cumulative time, cumulative water volume, bulk density and saturated hydraulic conductivity were used as model inputs. A standard shuffled cross-validation approach was applied and model performance was assessed using statistical indices including R^2 , RMSE, and MAE.

The results indicated that the most accurate estimates of infiltration depth ($R^2 = 0.94$, RMSE = 0.27 cm, MAE = 0.10 cm) and surface width ($R^2 = 0.98$, RMSE = 1.58 cm, MAE = 0.91 cm) were obtained for the Sandy Clay soil using the CatBoost model. The weakest performance was observed for the Elnet model. Specifically, the lowest accuracy for estimating infiltration depth and surface width using this model was obtained in Loamy Sand ($R^2 = 0.16$, RMSE = 6.16 cm, MAE = 3.22 cm) and Sandy Clay ($R^2 = 0.76$, RMSE = 6.09 cm, MAE = 4.67 cm), respectively.

Overall, the findings confirm the strong capability of tree-based machine learning models which are able to capture complex and nonlinear relationships among variables in estimating the dimensions of wetting bulb, particularly in soils with higher clay content.

Keywords: CatBoost algorithm, Data-driven models, Wetting pattern dimensions, Simulation.

EXTENDED ABSTRACT

Introduction

The increasing strain on water resources, resulting from climate change and rising demand for food production, has posed significant challenges to the transition toward sustainable agriculture. Consequently, the use of traditional irrigation methods such as surface irrigation—which are often associated with low efficiency and substantial water loss—is no longer justifiable. This has led to a growing interest among farmers in pressurized irrigation systems, such as drip irrigation, due to its precise control over applied water volume and improved water use efficiency. However, the performance of drip irrigation systems is substantially enhanced when accurate information regarding the dimensions of the soil moisture distribution pattern is available. Numerous analytical, numerical, and empirical models have been developed to estimate the infiltration depth and surface wetted width of the moisture distribution pattern. Nevertheless, these models have certain limitations, including the requirement for a wide range of input variables, the definition of complex boundary conditions, or the need for recalibration. With the emergence and advancement of machine learning models, robust algorithms have been introduced that can provide accurate estimations of the moisture distribution pattern dimensions without the aforementioned constraints. Therefore, the present study evaluates the accuracy of five machine learning models for estimating the infiltration depth and surface wetted width of the soil moisture distribution pattern over time under saline water application in surface drip irrigation. As the first step, a laboratory setup was constructed to record the aforementioned characteristics.

Method

a) Construction of the Laboratory Setup

To investigate the soil moisture distribution pattern, a physical model was constructed in the Irrigation Laboratory of Shahid Bahonar University of Kerman. The setup consisted of several components including a water supply source, a soil reservoir, and a surface drip irrigation system. The physical model comprised three separate compartments, with dimensions of $0.8 \times 0.8 \times 1$ m. A transparent Plexiglas sheet was installed on the front side of each compartment, gridded into 5×5 cm squares to facilitate visual observation and measurement. Each compartment was filled layer by layer with soil samples passed through a 2 mm sieve, and each layer was compacted using 15 blows of a 5 kg weight. A surface drip emitter with 4 L/h discharge rate was placed on the soil surface near the transparent Plexiglas wall so that the infiltration depth and surface wetted width could be clearly observed and measured. Five salinity treatments with electrical conductivities (EC) of 1.5, 3, 6, 9, and 12 $\text{dS} \cdot \text{m}^{-1}$ were applied to examine the moisture distribution pattern. The infiltration depth and surface wetted width were recorded at 10 time intervals of 30 minutes, with three replications. The collected data from two soil textures were used to simulate the dimensions of the moisture distribution pattern and to evaluate the accuracy of selected machine learning models.

b) Machine Learning Models

RF: Random Forest is a supervised ensemble learning model that consists of a large number of decision trees. The model is developed using the bagging technique combined with random feature selection at each splitting node. This process generates multiple independent decision trees, and the final output is determined by majority voting in classification problems or by averaging the predictions in regression tasks.

XGBoost: eXtreme Gradient Boosting is an efficient and scalable implementation of the Gradient Boosting framework based on decision trees. The algorithm builds trees in a sequential (additive) manner, where each successive tree is trained to correct the residual errors of the preceding ones. With using advanced regularization, gradient-based optimization and parallel processing, XGBoost delivers high accuracy and strong generalization performance.

CatBoost: Categorical Boosting is a gradient boosting algorithm built on decision trees that is specifically engineered for the direct handling of categorical features without requiring preprocessing. The algorithm utilizes Ordered Boosting and Ordered Target Statistics to minimize target leakage, control overfitting, and deliver superior performance in both accuracy and training speed.

SVR: Support Vector Regression (SVR) is a supervised learning algorithm developed for regression problems. It aims to find a function that deviates from the actual observed values by no more than a specified margin ϵ , while maintaining the flattest possible function. For nonlinear relationships, SVR employs kernel functions (such as the Radial Basis Function - RBF kernel) to implicitly map the input data into a higher-dimensional feature space. This transformation allows the algorithm to perform linear regression in the new space, which corresponds to nonlinear regression in the original input space.

Elnet: Elastic Net is a regularized linear regression model that integrates the L1 (Lasso) and L2 (Ridge) penalties. The model minimizes an objective function consisting of the residual sum of squares augmented by a convex combination of L1 and L2 regularization terms. This hybrid regularization enables both automatic feature selection (sparsity) and the handling of multicollinearity through the grouping effect, making Elastic Net highly effective for high-dimensional data and datasets with strongly correlated predictors.

c) Implementation and Evaluation of the Models

The implementation of the aforementioned machine learning models was carried out in Python 3.13. The input variables included sand percentage, silt percentage, irrigation water salinity, cumulative time, cumulative water volume, bulk density, and saturated hydraulic conductivity. These inputs were used to predict the surface wetted width (width) and the infiltration depth (depth) of the soil moisture distribution pattern.

The dataset was normalized using the min-max normalization method. To improve model generalization and prevent data leakage, shuffled stratified k-fold cross-validation was employed. The performance of the models was evaluated using the statistical indices coefficient of determination (R^2), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE).

Results

The analysis of the obtained results confirms the acceptable accuracy of the tree-based models in estimating the depth and surface width of the soil moisture distribution pattern in both soil textures. The most accurate estimation of depth was achieved by the CatBoost model in sandy clay soil ($R^2 = 0.94$, RMSE = 0.27 cm, MAE = 0.10 cm) and in loamy sand soil ($R^2 = 0.92$, RMSE = 1.62 cm, MAE = 0.77 cm). The weakest performance in estimating depth in both soil textures was observed with the Elnet model. Similarly, the most accurate estimation of surface width was obtained by the CatBoost model in sandy clay ($R^2 = 0.98$, RMSE = 1.58 cm, MAE = 0.91 cm) and in loamy sand

($R^2 = 0.97$, RMSE = 1.76 cm, MAE = 1.06 cm). The weakest results for width prediction in both textures were also recorded by the Elnet model.

Conclusions

In the present study, the performance of five machine learning models including CatBoost, XGBoost, RF, SVR and Elnet was evaluated for estimating the infiltration depth and surface width of the soil moisture distribution pattern in two different soil textures under surface drip irrigation. The results indicated that in both soil textures, tree-based models exhibited higher accuracy compared to the other models in estimating the aforementioned variables. Among them, CatBoost provided the most accurate estimates of both infiltration depth and surface width across both soil textures, often showing close agreement with the results obtained from the XGBoost model. This finding suggests that these models are more capable of capturing complex and hidden patterns among the variables.

In contrast, the presence of nonlinear relationships between input and target variables in simulating soil moisture distribution patterns, along with significant interactions among variables such as water volume, irrigation duration and soil properties led to weaker performance in models such as Elnet regression, particularly in soils with high sand content.

Furthermore, comparison of the results revealed that in soils with higher clay content, water movement in both vertical (depth) and horizontal (surface width) directions is more regulated and system variations occur in a smoother and more continuous manner. Consequently, the relationships among variables are more structured, resulting in more stable model performance for both target variables.

پیش‌نویس
دی‌نشر

مقدمه

کشاورزی به عنوان یکی از قدیمی‌ترین و مهم‌ترین فعالیت‌های بشری همواره با مجموعه‌ای از انتخاب‌ها، پیچیدگی‌ها و چالش‌هایی مانند کمبود منابع آبی به‌ویژه در مناطق خشک و نیمه خشک مواجه بوده است (Kusumavathi et al., 2025). عواملی مانند افزایش جمعیت و تقاضا برای غذا، تغییر الگوی مصرف، وقوع بلایای طبیعی، خشک‌سالی و کاهش اراضی قابل کشت منجر به استفاده بیش

از حد از نهاده‌هایی مانند آب در بخش کشاورزی شده که نتیجه آن تخلیه سریع منابع آبی و همچنین ایجاد آلودگی‌های زیست محیطی است (Sishodia et al., 2020). این عوامل در کنار رقابت بین بخش‌های مختلف در تخصیص آب باعث شده تا در بخش کشاورزی به تکنیک‌ها و روش‌هایی توجه شود که زمینه‌ساز استفاده بهینه از آب و حرکت به سمت کشاورزی پایدار باشند (Delgado et al., 2019; Sishodia et al., 2020). از این رو سیستم آبیاری قطره‌ای به دلیل داشتن مزایایی مانند جلوگیری از ایجاد رواناب، کاهش نفوذ عمقی، کاهش تبخیر و بهبود بهره‌وری مصرف آب به عنوان یک گزینه مناسب جهت استفاده کارآمد از منابع آبی مورد توجه قرار گرفته و این پتانسیل‌ها در شرایطی عملیاتی خواهند شد که جنبه‌های طراحی و اجرا و مدیریت سیستم به صورت بهینه اعمال شود (Cristóbal et al., 2015). از آنجا که ابعاد پیاز رطوبتی تحت تاثیر فاکتورهای متعددی مانند خصوصیات خاک، رطوبت اولیه، دبی قطره‌چکان، تعداد و فاصله بین قطره‌چکان‌ها، مدیریت آبیاری و ویژگی‌های ریشه گیاه و تبخیر قرار دارد، مدل‌های مختلفی برای شبیه‌سازی آن ارائه شده است که شامل مدل‌های تحلیلی، عددی و تجربی می‌باشند. مدل‌های تجربی بر مبنای روابط رگرسیونی حاکم بر داده‌های آزمایشگاهی یا مزرعه‌ای توسعه داده شده‌اند (کاندلوس و همکاران، ۱۳۸۷; Al-Ogaidi et al., 2015; Al-Ogaidi et al., 2011; Malek & Peters 2011; Cook et al., 2003; Kandelous & Šimůnek 2010; Hammami & Zayani 2016; Moncef & Khemaies 2016) و مدل‌های تحلیلی (Elmaloglou et al., 2013; Arbat et al., 2013; Sejna et al., 2014) بر مبنای حل معادلات ریچاردز تحت شرایط مرزی و اولیه خاص ارائه شده‌اند. استفاده از مدل‌های تحلیلی و عددی نیاز به مهارت در کاربرد و در دسترس بودن اطلاعات کافی دارد و مدل‌های تجربی نیز بر خلاف سادگی‌شان عمومی نیستند و صرفاً برای شرایط مشخص تعریف شده‌اند.

بنابراین ارائه یک رویکرد نوآورانه که از انعطاف‌پذیری و دقت بالا و نیاز به حداقل داده ورودی بهره‌مند باشد، اقدام موثری جهت مطالعه چگونگی توزیع آب در خاک بوده و در این راستا مدل‌های یادگیری ماشین می‌توانند با تحلیل هوشمندانه داده‌ها، ابعاد پیاز رطوبتی را با دقت مناسب برآورد کنند و جایگزین مطمئنی برای سایر مدل‌های ذکر شده باشند. در واقع مدل‌های یادگیری ماشین که به مدل‌های داده‌محور نیز معروفند، برای استخراج الگوهای رفتاری با استفاده از آموزش توسعه یافته‌اند. این مدل‌ها قادرند روابط پیچیده و غیرخطی بین متغیرهای ورودی و خروجی را شناسایی و ویژگی‌ها یا الگوهای معنادار موجود در داده‌ها را تشخیص دهند. به علاوه این مدل‌ها به دلیل بهره‌گیری از راهبردهای پیشرفته یادگیری، می‌توانند با مجموعه داده‌هایی که دارای توزیع احتمال ناشناخته هستند، کار کنند و امکان ترکیب چندین مدل با استفاده از روش‌های تجمیعی را نیز فراهم کند تا دقت و پایداری کلی پیش‌بینی‌ها افزایش یابد (Taheri et al., 2025).

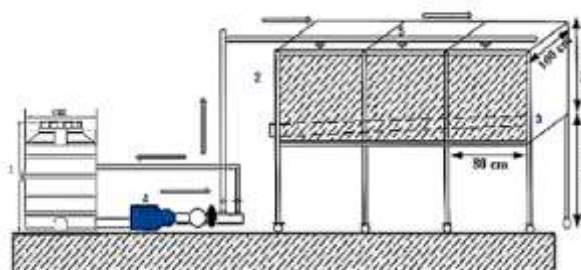
مرور پژوهش‌های مرتبط با توزیع رطوبت در خاک نشان می‌دهد که بخش عمده تحقیقات انجام شده بر ارزیابی دقت مدل‌های یادگیری ماشین در تخمین میزان رطوبت خاک متمرکز است و مطالعات محدودتری در زمینه تخمین ابعاد پیاز رطوبتی انجام شده (Amin & Ekhmaj 2006; Samadianfard et al., 2012; Karimi et al., 2020; Shiri et al., 2020) که ممکن است به دلیل ماهیت مدل‌ها، پویایی آب در خاک و یا عدم دسترسی به اطلاعات کافی باشد. اما پیشرفت در تکنیک‌های تقویت آموزش، منجر به ظهور الگوریتم‌های مقاومتری شده است که با حداقل داده و حتی در شرایط عدم وجود تنوع بین متغیرهای ورودی نیز قادر به ارائه نتایج قابل قبول هستند و البته ضروری است که عملکرد آنها در شرایط مختلف مورد بررسی قرار گیرد تا قابلیت تعمیم‌پذیری و اطمینان عملی آن‌ها تأیید شود.

از آنجا که پویایی توزیع رطوبت در خاک بر عرض سطحی و عمق نفوذ پیاز رطوبتی تاثیرگذار است، تخمین دقیق این ابعاد از مهم‌ترین نیازهای طراحی و مدیریت سیستم‌های آبیاری قطره‌ای محسوب شده و نقش تعیین‌کننده‌ای در بهینه‌سازی عملکرد سیستم، مدیریت آب و انرژی و افزایش کارایی مصرف آب دارد. لذا در تحقیق حاضر به ارزیابی دقت برخی از مدل‌های یادگیری ماشین جهت تخمین ابعاد پیاز رطوبتی پرداخته شده و در اولین گام، ساخت یک مدل آزمایشگاهی برای ثبت صفات عمق نفوذ و عرض سطحی در بازه‌های زمانی مختلف تحت آبیاری قطره‌ای سطحی انجام شده است.

مواد و روش ها

الف) ساخت مدل آزمایشگاهی

به منظور بررسی ابعاد پیاز رطوبتی، ابتدا یک مدل فیزیکی در آزمایشگاه آبیاری دانشگاه شهید باهنر کرمان ساخته شد که شامل بخش‌های مختلف از جمله منبع تامین آب، مخزن خاک و سیستم آبیاری قطره‌ای سطحی می‌باشد که نمای کلی آن در شکل (۱) ارائه شده است. برای ساخت مدل، یک سازه فلزی به ابعاد $1 \times 1 \times 2/4$ ساخته شد که به سه بخش مساوی تقسیم شد. در قسمت جلویی بخش‌های سه‌گانه و به منظور مشاهده الگوی توزیع رطوبت و ثبت اطلاعات یک ورق پلکسی گلس شفاف به ابعاد 80×80 سانتی‌متر که به اندازه 5×5 سانتی‌متر شطرنجی شده بود، نصب و آب‌بندی شد و برای جلوگیری از ایجاد جریان ترجیحی در مرز تماس خاک و دیواره شفاف، از یک لایه نازک گریس روی صفحه شفاف استفاده شد (سعیدی عباس آباد و همکاران، ۱۴۰۳).



شکل ۱. نمای کلی مدل آزمایشگاهی ساخته شده شامل ۱- مخزن آب ۲- مخزن خاک ۳- لوله زهکش ۴- پمپ آب ۵- قطره چکان

برای پر کردن هر بخش مدل، ابتدا قلوه سنگ‌ها و بخش‌های گیاهی از نمونه خاک جدا شدند. سپس نمونه‌ها هوا خشک شد و از الک ۲ میلی‌متری عبور داده شد. برای ایجاد تراکم یکسان، خاک به صورت لایه به لایه با ضخامت ۱۰ سانتی‌متر به درون هر یک از بخش‌ها منتقل و هر لایه با ۱۵ ضربه یک وزنه ۵ کیلوگرمی کوبیده شد تا شرایط یکسان ایجاد گردد (نیکبخت و عبدالهی سیاهکلرودی، ۱۳۹۳). در این تحقیق از اطلاعات عمق نفوذ و عرض سطحی پیاز رطوبتی که در دو بافت Loamy Sand و Sandy Clay ثبت شده بود، استفاده گردید که برخی از مشخصات بافت خاک در جدول (۱) ارائه شده است.

جدول ۱. برخی از ویژگی‌های خاک مورد استفاده

جرم مخصوص ظاهری (gr/cm^3)	Ks (cm/day)	pH	EC (dS/m)	کلاس بافت	سیلت (%)	رس (%)	ماسه (%)
۱/۵۵	۷۷/۰۸	۸/۱۷	۴/۸۳	Loamy sand	۸	۱۱	۸۱
۱/۴۲	۱۳/۴۱	۸/۲۶	۴/۵۲	Sandy clay	۱۰	۳۸	۵۲

برای تامین آب در مدل از یک سیستم آبیاری قطره‌ای سطحی با قطره‌چکان ۴ لیتر بر ساعت و فشار کاری ۱ اتمسفر استفاده شد. قطره چکان مورد استفاده در این تحقیق از مدل ژئوفلو که از نوع خود شوینده و جبران کننده فشار است، انتخاب شد و هر قطره‌چکان روی سطح خاک و در مجاورت صفحه شفاف قرار گرفت تا ابعاد پیاز رطوبتی قابل رویت و اندازه‌گیری باشد. در این تحقیق از ۵ تیمار

مختلف آب با سطوح شوری ($ds.m^{-1}$) ۱۲، ۹، ۶، ۳، ۱/۵) استفاده شد که برای ساخت تیمارها از کلرید سدیم و روابط (۱-۲) استفاده گردید. همچنین در هر مرحله برای اطمینان از تامین سطح شوری مورد نظر، کنترل سطوح با استفاده از دستگاه EC سنج انجام شد.

$$TDS = 640 \times EC \quad EC < 5 \text{ ds.m}^{-1} \quad \text{رابطه ۱}$$

$$TDS = 800 \times EC \quad EC \geq 5 \text{ ds.m}^{-1} \quad \text{رابطه ۲}$$

در روابط فوق EC شوری آب ($ds.m^{-1}$) و TDS غلظت کلرید سدیم مورد استفاده ($mgr.lit^{-1}$) است (Hanson et al., 2006). آب شور از طریق قطره‌چکان روی سطح خاک در هر بخش از مدل توزیع شده و با اشباع شدن لایه سطحی، حرکت افقی و عمودی رطوبت از طریق صفحه شطرنجی شفاف قابل مشاهده بود و ابعاد پیاز رطوبتی شامل عرض سطحی و عمق نفوذ در ۱۰ بازه زمانی نیم ساعته و در سه تکرار ثبت شد. بعد از انجام هر آزمایش، خاک هر بخش از مدل به صورت جداگانه تخلیه و به صورت یک لایه نازک روی پلاستیک و به مدت ۲۲ ساعت هوا خشک شد تا اثر رطوبت ذخیره شده در خاک برای آزمایشات بعدی حذف شود و سپس با رعایت تراکم، مجدداً در مدل آزمایشگاهی قرار داده شد. همچنین از چندین مرتبه آبخوئی جهت حذف اثر تجمع نمک در خاک استفاده شد. از مجموعه اطلاعات جمع‌آوری شده برای شبیه‌سازی ابعاد پیاز رطوبتی و ارزیابی دقت برخی از مدل‌های یادگیری ماشین استفاده گردید.

ب) معرفی مدل‌های یادگیری ماشین جنگل تصادفی (RF)

جنگل تصادفی یک روش یادگیری ترکیبی است که در آن چندین درخت تصمیم با استفاده از یک زیرمجموعه که به طور تصادفی از نمونه‌هایی که از طریق جایگزینی (تکرار) تولید شده‌اند، ساخته می‌شود. به عبارت دیگر در مدل جنگل تصادفی که یک روش گروهی است، چندین مدل درختی برای تولید یک پیش‌بینی از هر پدیده ترکیب می‌شوند. همچنین مدل جنگل تصادفی این ویژگی را دارد که به دلیل تصادفی و ناهمبسته بودن، ارتباط بین متغیرهای ورودی و هدف را حتی در شرایطی که روابط آنها پیچیده و غیرخطی است، برقرار کند (Hong et al., 2016). الگوریتم رگرسیونی جنگل تصادفی شامل مراحل زیر است: ۱- نمونه‌های مختلف بوت‌استرپ X_i (i = تکرار بوت‌استرپ) به طور تصادفی از مجموعه داده اصلی X انتخاب می‌شوند. دو سوم از نمونه‌ها در یک نمونه بوت‌استرپ و یک سوم از نمونه‌ها خارج از کیسه قرار داده می‌شوند. هر درخت برای مطابقت با یک زیرمجموعه خاص از بوت‌استرپ ساخته می‌شوند. ۲- در هر گره از هر درخت، ابتدا یک زیرمجموعه تصادفی از ویژگی‌ها انتخاب می‌شود. سپس بهترین ویژگی و نقطه تقسیم با کمترین MSE از میان همین زیرمجموعه انتخاب می‌گردد. ۳- فرایند تقسیم داده‌ها در هر گره داخلی مطابق با گام‌های بالا تا زمانی که تمام درختان تصادفی رشد کنند و شرایط توقف حاصل گردد، تکرار می‌شود. نتایج نهایی رگرسیون می‌تواند بر اساس رابطه (۳) محاسبه شود:

$$\hat{y}(X_i) = \frac{1}{K} \sum_{k=1}^K T_b(X_i) \quad \text{رابطه ۳}$$

در رابطه فوق K تعداد درختان جنگل و T_b نشان دهنده هر درخت است (Wang et al., 2018).

الگوریتم تقویت گرادیانی پیشرفته XGBoost

الگوریتم XGBoost یک پیاده‌سازی بهینه‌شده و مقیاس‌پذیر از روش درخت‌های تصمیم تقویتی مبتنی بر گرادیان است که در چارچوب گرادیان بوستینگ عمل می‌کند و با بهره‌گیری از تکنیک‌هایی مانند منظم‌سازی، هرس درخت و محاسبات موازی، دقت پیش‌بینی و

¹ - Random Forest

² - eXtreme Gradient Boosting

قابلیت تعمیم‌پذیری و سرعت آموزش را افزایش می‌دهد (Chen & Guestrin 2016). این مدل مشابه سایر روش‌های گرادیان بوستینگ، دقت پیش‌بینی‌ها را از طریق کمینه‌سازی تابع زیان باقیمانده‌ها افزایش می‌دهد. از آنجا که این مدل فقط از درخت‌های تصمیم به عنوان مدل پایه استفاده می‌کند، نسخه اصلاح شده‌ای از تابع زیان را جهت کنترل پیچیدگی درخت‌ها به کار می‌برد و این کنترل پیچیدگی، منجر به افزایش سرعت آموزش و کاهش فضای ذخیره می‌گردد. از ویژگی‌های این مدل می‌توان به استفاده از الگوریتم آگاه از پراکندگی برای یافتن بهترین تقسیم، یادگیری گره پیش‌فرض برای هدایت و حذف نمونه‌های صفر، قیود یکنواختی که مدل را مجبور می‌کند نسبت به برخی ویژگی‌ها فقط در جهت افزایشی یا کاهش‌ی تغییر کند و قیود تعامل ویژگی‌ها که ترکیب ویژگی‌هایی را که می‌توانند در مسیر ریشه تا برگ قرار گیرند محدود می‌کند، اشاره کرد (Bentéjac et al., 2021).

در الگوریتم XGBoost تابع ضرر معمولاً به صورت جمع کل اختلاف بین مقادیر حاصل از پیش‌بینی مدل و مقادیر واقعی با روابط زیر محاسبه می‌شود.

$$(y_i, \tilde{y}_i) = (y_i - \tilde{y}_i)^2 \quad \text{رابطه ۴}$$

$$\Omega(f) = \gamma \times T + \frac{1}{2} \lambda \times \sum_{j=1}^T w_j^2 \quad \text{رابطه ۵}$$

در روابط مذکور γ ماتریس مقادیر واقعی، \tilde{y}_i ماتریس مقادیر پیش‌بینی شده، T تعداد گره‌ها، w_j وزن گره‌ها، λ و γ پارامترهای منظم‌سازی هستند (Dhaliwal et al., 2022).

الگوریتم CatBoost

CatBoost یک مدل درخت تصمیم مبتنی بر گرادیان بوستینگ است که هدف اصلی آن کاهش خطاهایی است که ناشی از تفاوت رفتار مدل بین داده‌های آموزشی و داده‌های تست می‌باشد و دلیل این اختلاف، آن است که در روش‌های معمول گرادیان بوستینگ از یک مجموعه داده هم برای محاسبه گرادیان و هم برای ساخت درخت‌های افزایشی استفاده می‌شود. در نتیجه مدل ممکن است به شکلی نادرست یاد بگیرد و روی داده‌های آموزشی بهتر از داده‌های تست عمل کند. CatBoost برای حل این مشکل روش Ordered Boosting را ارائه می‌دهد که در آن گرادیان هر نمونه با استفاده از مدل‌هایی تخمین زده می‌شود که این نمونه در آموزش وجود نداشته است. بنابراین به جای ساخت صدها مدل مختلف، تنها یک مدل در هر تکرار ساخته می‌شود که تمام این ترتیب‌های تصادفی و مدل‌ها را مدیریت می‌کند. مدل‌های پایه در CatBoost درخت‌های متقارن هستند، یعنی همه گره‌های یک سطح با یک شرط تقسیم یکسان گسترش می‌یابند و مدل نهایی به صورت جمع سریالی درخت‌ها ساخته می‌شود (Prokhorenkova et al., 2018).

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i) + c \quad \text{رابطه ۶}$$

که در این رابطه f_k درخت k ام و C مقدار ثابت اولیه است.

رگرسیون E_{net}

این رگرسیون که برای نخستین بار توسط (Zou & Hastie 2005) معرفی شد، یک روش منظم‌سازی برای مدل‌های رگرسیون است که جهت رفع مشکلات ناشی از هم‌خطی چندگانه و ابعاد بالای داده‌ها توسعه یافته است. این روش ترکیبی از جریمه‌های (L1) لاسو^۶

³ - Interaction Constraints

⁴ - Categorical Boosting

⁵ - Elastic net

⁶ - Lasso

و (L2) راجع⁷ بوده و به طور هم‌زمان قابلیت انتخاب متغیر و پایداری ضرایب برآوردی را فراهم می‌کند. اگر مجموعه‌ای از داده شامل n مشاهده و p متغیر مستقل باشد، مدل رگرسیون خطی به صورت رابطه⁸ (7) تعریف می‌شود:

$$y_i = \beta_0 + \sum_{j=1}^p x_{ij} \beta_j + \varepsilon_i, i = 1, \dots, n \quad (\text{رابطه 7})$$

که در آن y_i متغیر پاسخ، x_{ij} متغیرهای توضیحی، β_j ضرایب رگرسیون و ε_i جمله خطا است. در مدل الاستیکنت، ضرایب رگرسیون از طریق کمینه‌سازی تابع هدف زیر برآورد می‌شوند:

$$\hat{\beta}_{EN} = \arg \min_{\beta} \left\{ \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j \right)^2 + \lambda_1 \sum_{j=1}^p |\beta_j| + \lambda_2 \sum_{j=1}^p \beta_j^2 \right\} \quad (\text{رابطه 8})$$

در رابطه فوق λ_1 پارامتر تنظیمی متناظر با جریمه L1 مرتبط با لاسو و λ_2 پارامتر تنظیمی متناظر با جریمه L2 مرتبط با راجع است. (Zou & Hastie 2005) نشان دادند که می‌توان این تابع جریمه را به صورت ترکیبی و با استفاده از پارامتر تنظیم‌کننده α به صورت رابطه (9) بازنویسی کرد:

$$\alpha = \frac{\lambda_1}{\lambda_1 + \lambda_2}, 0 \leq \alpha \leq 1 \quad (\text{رابطه 9})$$

به طوری که اگر $\alpha = 1$ باشد، مدل به رگرسیون لاسو کاهش می‌یابد. اگر $\alpha = 0$ باشد، مدل معادل رگرسیون راجع خواهد بود و برای $0 < \alpha < 1$ مدل الاستیکنت حاصل می‌شود. یکی از ویژگی‌های مهم رگرسیون الاستیکنت اثر گروهی است. بر اساس این ویژگی و در صورت وجود متغیرهای مستقل با همبستگی بالا، مدل تمایل دارد ضرایب این متغیرها را به طور هم‌زمان و با مقادیر مشابه برآورد کند. این درحالی است که رگرسیون لاسو معمولاً تنها یکی از متغیرهای همبسته را به صورت دلخواه انتخاب کرده و سایر متغیرها را حذف می‌کند. همچنین در شرایطی که تعداد متغیرها از تعداد مشاهدات بیشتر باشد ($p > n$)، رگرسیون الاستیکنت قادر است بیش از n متغیر را در مدل نگه دارد و در نتیجه، انعطاف‌پذیری بیشتری در مدل‌سازی داده‌های با ابعاد بالا دارد. لذا رگرسیون الاستیکنت به عنوان روشی کارآمد برای مدل‌سازی داده‌هایی با هم‌خطی چندگانه و تعداد زیاد متغیرهای توضیحی شناخته می‌شود و می‌تواند جایگزین مناسبی برای روش‌های کلاسیک رگرسیون در چنین شرایطی باشد (Zou & Hastie 2005).

رگرسیون ماشین بردار پشتیبان SVR

رگرسیون ماشین بردار پشتیبان توسط (Vapnik & Chervonenkis 1974; Vapnik 1984) بر پایه تئوری یادگیری آماری توسعه داده شد و یکی از روش‌های یادگیری نظارت‌شده است که از آن برای حل مسائل رگرسیون استفاده می‌شود. این مدل بر پایه اصل حداقل‌سازی ریسک ساختاری⁹ توسعه یافته است و تلاش می‌کند تعادل مناسبی بین پیچیدگی مدل و خطای آموزش برقرار کند. در حالی که بسیاری از شبکه‌های عصبی مصنوعی مانند پرسپترون چندلایه عمدتاً بر پایه اصل حداقل‌سازی ریسک تجربی¹⁰ عمل می‌کنند که تنها خطای آموزش روی داده‌های مشاهده‌شده را حداقل می‌نماید. در SVR هدف پیدا کردن تابع $f(x)$ است. به نحوی که برای همه

⁷ - Ridge

⁸ - Support Vector Regression

⁹ - Structural Risk Minimization

¹⁰ - Empirical Risk Minimization

داده‌های آموزشی حداکثر به میزان ϵ از هدف‌های بدست آمده واقعی y_i انحراف داشته و در عین حال تا حد ممکن هموار باشد که به صورت رابطه (۱۰) بیان می‌شود.

$$f(x) = w \cdot \phi(x) + b \quad \text{رابطه (۱۰)}$$

به گونه‌ای که $\epsilon \geq |f(x_i) - y_i|$ برای اکثر نقاط آموزشی برقرار باشد. در رابطه فوق $f(x)$ تابع پیش‌بینی، w بردار وزن در فضای ویژگی، $\phi(x)$ تابع نگاشت غیرخطی که داده‌های ورودی را به فضای ابعاد بالاتر تبدیل می‌کند و b مقدار بایاس می‌باشد. در واقع در این مدل در حالتی که نتوان یک تابع رگرسیون خطی در فضای ورودی به داده‌ها برازش داد، از یک نگاشت غیرخطی ϕ برای انتقال داده‌ها به یک فضا با ابعاد بالاتر استفاده می‌شود و سپس در این فضای جدید، رگرسیون خطی انجام می‌شود که معادل رگرسیون غیر خطی در فضای ورودی است. انجام محاسبات در فضای با ابعاد زیاد با محدودیت‌هایی مواجه است و برای غلبه بر این مشکل از تابع کرنل استفاده می‌شود که این تابع امکان محاسبه ضرب داخلی بردارها در فضای ویژگی را بدون نیاز به نگاشت صریح فراهم می‌کند. از مزایای تابع کرنل می‌توان به حل مسائل محاسباتی که دارای ابعاد زیادی هستند و کارآمد بودن از نظر زمانی و حافظه اشاره کرد (Smola & Schölkopf 2004).

ج) پیاده سازی مدل‌های مورد استفاده

اجرا و پیاده سازی مدل‌های ذکر شده در محیط نرم افزار Python 3.13 انجام شد. در تحقیق حاضر از اطلاعات هشت پارامتر شامل درصد شن، درصد سیلت، درصد رس، شوری آب، زمان تجمع، حجم آب تجمع، وزن مخصوص ظاهری و هدایت هدرولیکی اشباع به عنوان متغیر ورودی به مدل‌ها استفاده شد و تخمین صفات عرض سطحی (width) و عمق نفوذ (depth) پیاز رطوبتی انجام شد. به علت تعدد متغیرهای ورودی و واحدهای اندازه‌گیری و جهت تسهیل در مقایسه داده‌ها، نرمال کردن داده‌ها قبل از آموزش مدل با استفاده از رابطه (۱۱) انجام شد:

$$X_n = \frac{X_i - X_{min}}{X_{max} - X_{min}} \quad \text{رابطه (۱۱)}$$

در رابطه فوق X_n داده استاندارد شده، X_i داده مشاهداتی و X_{min} و X_{max} به ترتیب بیشترین و کمترین داده‌ها است. همچنین در تحقیق حاضر از روش اعتبارسنجی متقاطع استاندارد با اختلاط تصادفی (برای افزایش دقت تخمین استفاده شد. در این روش ابتدا داده‌ها به صورت تصادفی درهم ریخته می‌شوند و سپس به k زیرمجموعه مستقل تقسیم می‌شوند. مدل در k تکرار روی $k-1$ بخش آموزش داده شده و روی بخش باقیمانده تست می‌شود و نتیجه نهایی با میانگین‌گیری معیارهای ارزیابی در تمام تکرارهای اعتبارسنجی آمحا سبه می‌شود. (Vanwinckelen & Blockeel 2012). این روش با بهره‌گیری مؤثر از کل داده‌ها، موجب کاهش واریانس برآورد و ارائه ارزیابی قابل اعتمادتر از توانایی تعمیم مدل بر روی داده‌های جدید می‌شود. در تحقیق حاضر به منظور یکسان سازی شرایط آموزش و دستیابی به نتایج قابل قبول، از روش اعتبارسنجی متقاطع پنج‌بخشی^۱ استفاده شد که در آن داده‌ها به پنج بخش تقریباً مساوی تقسیم شدند و در هر تکرار، چهار بخش (معادل ۸۰٪ داده‌ها) برای آموزش مدل و یک بخش (معادل ۲۰٪ داده‌ها) برای آزمون مورد استفاده قرار گرفت. مهم‌ترین ویژگی مدل‌های مورد مطالعه در جدول (۲) ارائه شده است.

^۱ - shuffled standard k-fold cross-validation

^۱ - fold

^۱ - 5 fold cross-validation

جدول ۲. ویژگی مدل‌های مورد مطالعه

ML Models	Optimal Parameters
RF	Bootstrap:True, max_depth:None, max_features:1, min_samples_leaf:1, min_samples_split: 2, n_estimators: 200, random state:42, n jobs:-1
SVR	Kernel:rbf, C:100, gamma:0.1, epsilon:0.01
XGBoost	subsample: 0.9, reg_lambda:1, reg_alpha:0.0, n_estimators:300, max_depth:6, learning_rate: 0.05, colsample_bytree: 0.9
CatBoost	subsample:None, learning_rate:0.1, leaf_reg:3, iterations:500, depth:5, random state=42, bootstrap type=Bayesian
Elnet	base_model = ElasticNet , alpha=0.05, l1_ratio=0.3, random_state=42, max_iter=20000, tol=1e-4, warm_start=True

د) ارزیابی عملکرد و دقت مدل‌ها

به منظور ارزیابی عملکرد مدل‌های ذکر شده از شاخص‌های آماری ضریب تعیین (R^2)، جذر میانگین مربعات خطا (RMSE) و میانگین خطای مطلق (MAE) به شرح زیر استفاده شد (Segovia et al., 2023).

$$R^2 = 1 - \frac{\sum_{i=1}^n (O_i - P_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2} \quad \text{رابطه ۱۲}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (P_i - O_i)^2}{n}} \quad \text{رابطه ۱۳}$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |P_i - O_i| \quad \text{رابطه ۱۴}$$

در روابط (۱۲-۱۴) مقدار پیش‌بینی شده توسط مدل O_i مقدار مشاهداتی، \bar{O} میانگین مقادیر مشاهداتی و n تعداد داده است. هر چه مقدار RMSE و MAE کمتر و به صفر نزدیک و مقدار R^2 به ۱ نزدیک باشد، نشان‌دهنده دقیق‌تر بودن شبیه‌سازی و کارایی بهتر مدل است.

نتایج و بحث

الف) عمق الگوی توزیع رطوبت در خاک

نتایج حاصل از محاسبه میانگین و انحراف معیار شاخص‌های آماری حاصل از تکرار اعتبارسنجی برای مدل‌های مورد بررسی در بافت Loamy Sand در دو مرحله آزمون و تست جهت تخمین عمق نفوذ پیاز رطوبتی در جدول (۳) ارائه شده است. لازم به ذکر است در این تحقیق برای هر بافت خاک از ۱۵۰ ردیف داده استفاده شد که بنا بر ماهیت تحقیق متغیرهایی مانند درصد رطوبت و شن و سیلت و وزن مخصوص ظاهری و... ثابت هستند. در واقع داده‌ها حاصل یک سری آزمایشات تکرار شونده با شرایط اولیه تقریباً یکسان هستند که در آنها فرآیند آبیاری از زمان صفر شروع و تا ۳۰۰ دقیقه ادامه پیدا می‌کند. هر چند این ساختار داده‌ای برای تحلیل رفتار نفوذ آب مناسب است، اما در شرایط کاربرد مدل‌های یادگیری ماشین با مشکل عدم تنوع در متغیرهای ورودی مواجه می‌شود و در هر تکرار اطلاعات جدید اندکی به مدل اضافه می‌شود. لذا به دلیل حجم محدود داده با تنوع کم نسبت به پیچیدگی مدل‌هایی مانند CatBoost و XGBoost انتظار بروز بیش برآزش وجود داشت که مقایسه نتایج بدست آمده این مطلب را تایید می‌کند. با این حال استفاده از روش اعتبارسنجی متقاطع منجر به بهبود عملکرد مدل‌ها شد. بررسی نتایج بدست آمده نشان داد در بافت Loamy Sand در مرحله تست بیشترین دقت بین تکرارهای اعتبارسنجی در مدل CatBoost با مقادیر $R^2 = 0.91 \pm 0.06$ ، $RMSE = 1.54 \pm 0.48$ cm و $MAE = 0.77 \pm 0.18$ cm

MAE و کمترین دقت در مدل رگرسیونی Elnet با مقادیر $R^2=0.16 \pm 0.038$ ، $RMSE = 5/86 \pm 1/90$ cm، $MAE=3/22 \pm 0/66$ cm مشاهده شد. همچنین بررسی شاخص‌های آماری محاسبه شده برای هر تکرار اعتبارسنجی نیز نشان داد که بهترین نتایج مرحله تست در تکرار اعتبارسنجی ۵ مدل CatBoost با مقادیر $R^2=0.92$ ، $RMSE=0/83$ cm، $MAE=0/56$ cm مشاهده شد. در مدل رگرسیونی Elnet نیز بهترین نتایج در مرحله تست در تکرار اعتبارسنجی ۵ با مقادیر $R^2=0.22$ ، $RMSE=6/43$ cm، $MAE=3/029$ cm مشاهده شد.

جدول ۳. شاخص‌های آماری محاسبه شده مدل‌های مذکور در مرحله آموزش و تست در بافت Loamy sand

ML Models	Depth					
	Train (Mean±SD Across Folds)			Test (Mean±SD Across Folds)		
	R ²	RMSE (cm)	MAE (cm)	R ²	RMSE (cm)	MAE (cm)
CatBoost	0/99±0/07	0/14±0/04	0/098±0/02	0/91±0/06	1/54±0/48	0/77±0/18
XGBoost	0/96±0/05	0/04±0/00	0/02±0/00	0/85±0/14	1/95±0/71	0/91±0/25
RF	0/81±0/04	1/39±0/08	0/70±0/05	0/64±0/24	3/19±0/61	1/79±0/28
SVR	0/28±0/04	5/69±0/58	1/81±0/19	0/31±0/09	5/38±2/30	1/86±0/76
Elnet	0/18±0/01	6/06±0/53	3/22±0/23	0/16±0/038	5/86±1/90	3/22±0/66

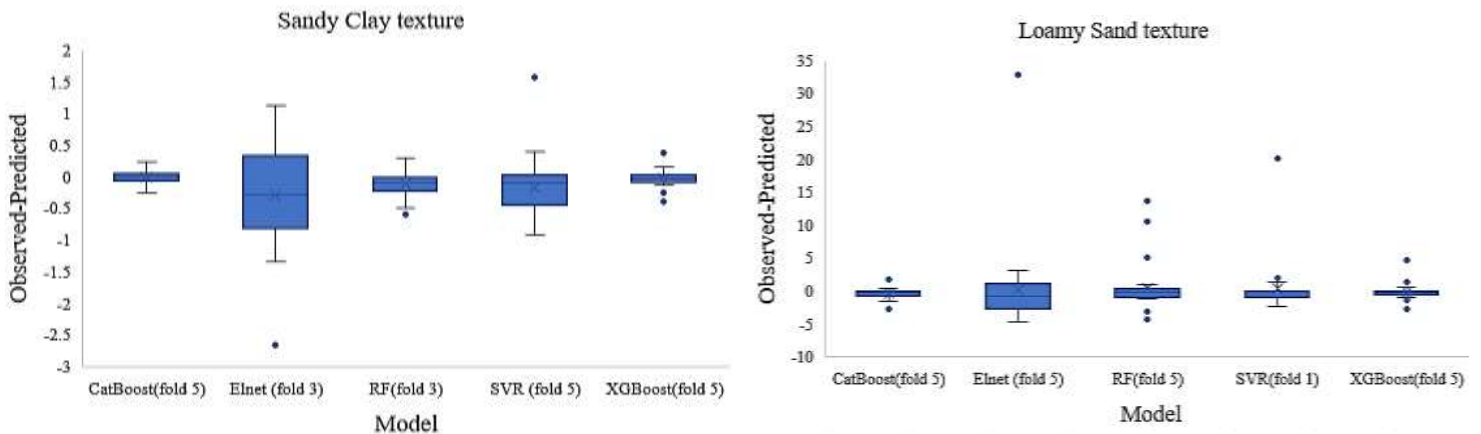
نتایج حاصل از محاسبه میانگین و انحراف معیار شاخص‌های آماری در تکرارهای اعتبارسنجی برای مدل‌های مورد بررسی در بافت Sandy Clay جهت تخمین عمق پیاز رطوبتی در جدول (۴) ارائه شده است. بررسی نتایج بدست آمده نشان داد که در این بافت بیشترین دقت در مرحله تست در مدل CatBoost با مقادیر $R^2=0.94 \pm 0.16$ ، $RMSE = 0/18 \pm 0/19$ cm، $MAE=0/10 \pm 0/07$ cm و کمترین دقت در مدل رگرسیونی Elnet با مقادیر $R^2=0.75 \pm 0/05$ ، $RMSE = 1/23 \pm 0/21$ cm، $MAE = 0/95 \pm 0/15$ cm مشاهده شد. همچنین بررسی شاخص‌های آماری محاسبه شده برای هر تکرار اعتبارسنجی نیز نشان داد که بیشترین دقت در مرحله تست در تکرار اعتبارسنجی ۳ مدل CatBoost با مقادیر $R^2=0.91$ ، $RMSE=0/52$ cm، $MAE = 0/36$ cm مشاهده شد و در مدل رگرسیونی Elnet نیز بهترین نتایج مرحله تست در تکرار اعتبارسنجی ۳ با مقادیر $R^2=0/82$ ، $RMSE=0/88$ cm، $MAE=0/70$ cm مشاهده شد.

جدول ۴. شاخص‌های آماری محاسبه شده مدل‌های مذکور در مرحله آموزش و تست در بافت Sandy Clay

ML Models	Depth					
	Train (Mean±SD Across Folds)			Test (Mean±SD Across Folds)		
	R ²	RMSE (cm)	MAE (cm)	R ²	RMSE (cm)	MAE (cm)
CatBoost	0/99±0/03	0/01±0/00	0/008±0/00	0/94±0/16	0/18±0/19	0/10±0/07
XGBoost	0/99±0/06	0/068±0/00	0/005±0/00	0/93±0/02	0/23±0/09	0/165±0/05
RF	0/98±0/03	0/16±0/01	0/11±0/00	0/91±0/03	0/40±0/14	0/28±0/09
SVR	0/92±0/00	0/67±0/03	0/43±0/03	0/87±0/03	0/68±0/17	0/47±0/13
Elnet	0/76±0/01	1/21±0/03	0/92±0/02	0/75±0/05	1/23±0/21	0/95±0/15

همچنین در شکل (۲) نمودار جعبه‌ای خطای پیش‌بینی (Observed - Predicted) جهت تخمین عمق نفوذ پیاز رطوبتی بر اساس نتایج اعتبارسنجی متقابل هر مدل در یک تکرار اعتبارسنجی منتخب ارائه شده است. تکرار اعتبارسنجی منتخب بر اساس بهترین مقادیر شاخص‌های آماری محاسبه شده در مرحله تست انتخاب شد و با مقایسات انجام شده مشخص گردید که در مجموع نزدیکی قابل توجهی با مقادیر سایر تکرارهای اعتبارسنجی نیز دارد. مقایسه نمودارهای جعبه‌ای رسم شده نشان داد که در بافت Loamy Sand مدل CatBoost دارای کمترین بازه میان‌چهارکی و کوتاه‌ترین خطوط فاصله‌ای است که بیانگر بیشترین پایداری تخمین و کمترین

میزان پراکندگی خطا می‌باشد و بیشترین خطا در مدل Elnet مشاهده شد. در بافت Sandy Clay نیز شباهت زیادی در پراکندگی مرکزی خطاها بین مدل CatBoost و XGBoost مشاهده شد، اما در مجموع دقت بالاتر مدل CatBoost به دلیل عدم وجود نقاط پرت و عملکرد یکنواخت‌تر تایید گردید.



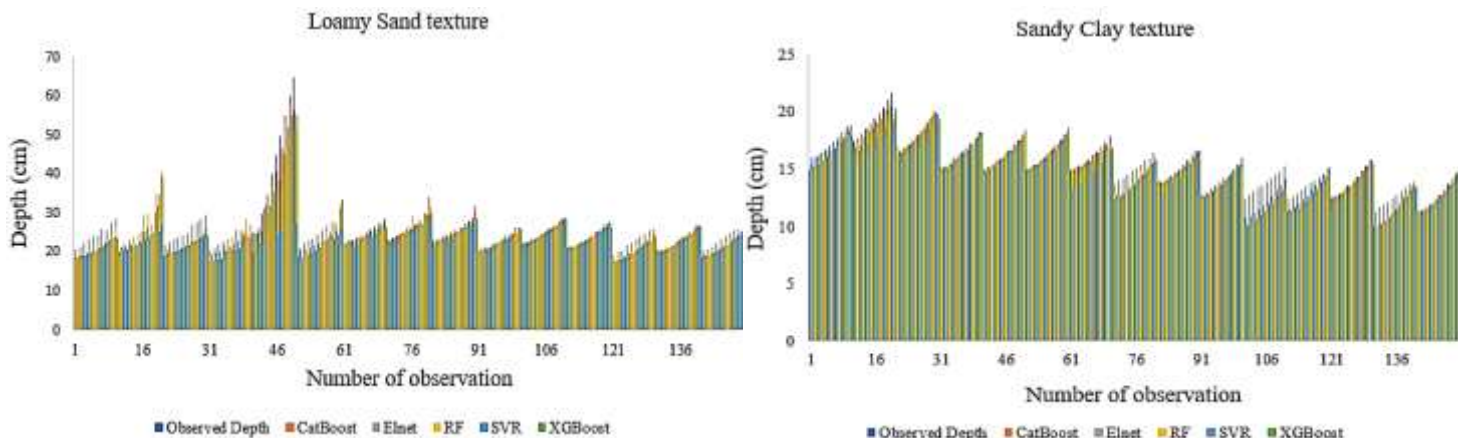
شکل ۲. نمودار جبهه‌ای خطای پیش‌بینی مدل‌های مورد بررسی در تخمین عمق نفوذ در پیاز رطوبتی

همچنین به منظور ارزیابی عملکرد جامع مدل‌ها از پیش‌بینی‌های خارج از تکرار اعتبارسنجی (OOB) نیز استفاده شد که نتایج آن در جدول (۵) ارائه شده و نمودارهای مقایسه عملکرد مدل‌های مورد بررسی و تیلور شاخص‌های آماری نیز رسم گردید (شکل ۳-۴). بررسی مقادیر مثبت شده نشان می‌دهد که در هر دو بافت مدل CatBoost بیشترین دقت و مدل RF کمترین دقت را در تخمین عمق نفوذ در بین مدل‌های درخت-پایه مورد بررسی داشته‌اند. همچنین مدل‌های SVR و Elnet از دقت کمتری در مقایسه با مدل‌های درخت-پایه در هر دو بافت برخوردار بودند و ضعیف‌ترین نتایج در مدل رگرسیونی Elnet و در بافت Loamy Sand مشاهده شد.

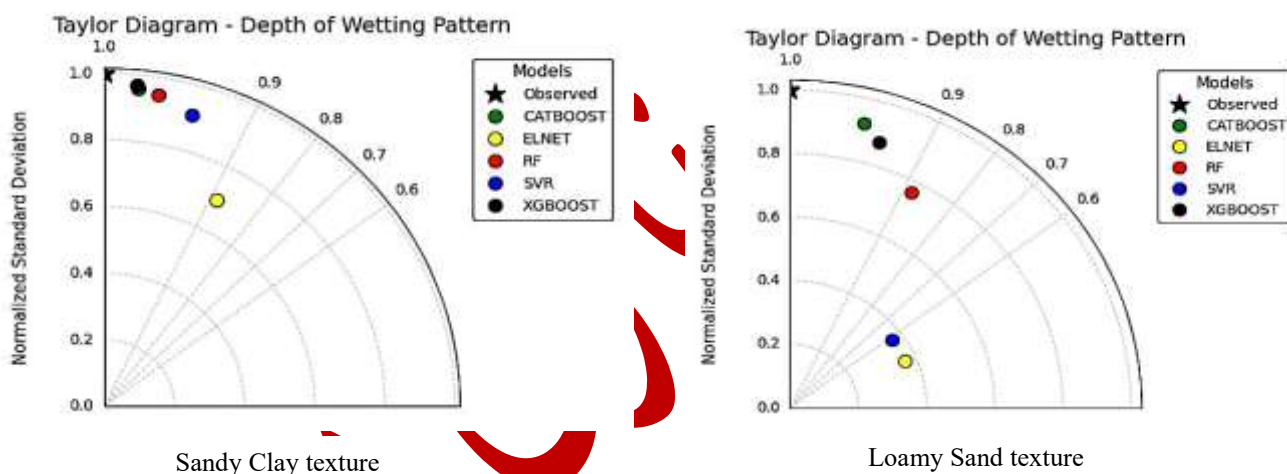
جدول ۵. عملکرد مدل‌های مورد بررسی در تخمین عمق نفوذ پیاز رطوبتی بر اساس پیش‌بینی‌های خارج از تکرار اعتبارسنجی (OOB)

ML \ بافت خاک	Sandy Clay			Loamy Sand		
	R ²	RMSE (cm)	MAE (cm)	R ²	RMSE (cm)	MAE (cm)
CatBoost	۰/۹۴	۰/۲۷	۰/۱۰	۰/۹۲	۱/۶۲	۰/۷۷
XGBoost	۰/۹۳	۰/۲۴	۰/۱۶	۰/۸۸	۲/۰۸	۰/۹۱
RF	۰/۹۲	۰/۴۲	۰/۲۸	۰/۶۹	۳/۲۴	۱/۷۹
SVR	۰/۸۸	۰/۷۰	۰/۴۷	۰/۲۶	۵/۷۶	۱/۸۶
Elnet	۰/۷۵	۱/۲۵	۰/۹۵	۰/۱۶	۶/۱۶	۳/۲۲

¹⁴ Out Of Fold



شکل ۳. مقایسه عملکرد مدل‌های مورد بررسی در پیش‌بینی عمق نفوذ پیاز رطوبتی



شکل ۴. نمودار تیلور عملکرد مدل‌های مورد بررسی در تخمین عمق نفوذ پیاز رطوبتی

در مجموع مقایسه نتایج حاصل نشان می‌دهد که دقت مدل‌ها در تخمین عمق نفوذ در پیاز رطوبتی تحت تاثیر دو عامل مهم است. عامل اول ویژگی مدل‌های مورد بررسی در مقایسه با حجم کم داده‌ها و عدم وجود تنوع بین متغیرهای ورودی است که ناشی از ماهیت چنین مطالعاتی به‌ویژه در شرایط آزمایشگاهی بوده و ناگزیر در هر تکرار، تفاوت اندکی بین داده‌ها مشاهده می‌شود که این مورد در مقایسه با پیچیدگی الگوریتم مدل‌هایی مانند CatBoost و XGBoost که ذاتاً برای شبیه‌سازی داده‌های پراکنده و متنوع ارائه شده‌اند باعث مقداری بیش‌برازش می‌شود که البته در محدوده قابل قبول قرار دارد. لازم به ذکر است که مدل‌های CatBoost و XGBoost مبتنی بر بوستینگ هستند و خصوصیتی مانند یادگیری روابط غیرخطی پیچیده، مدیریت بهتر تعاملات بین متغیرها، مقاومت در برابر نویز و کاهش بایاس از طریق یادگیری ترتیبی را دارند (Seifu Majdar et al., 2025). بنابراین از آنجا که حرکت آب در خاک‌های غیراشباع پیچیده است و توسط پیکربندی غیرخطی روابط بین ورودی‌ها (دبی، زمان، بافت خاک، رطوبت اولیه) کنترل می‌شود، این مدل‌ها بهتر می‌توانند الگوهای پنهان بین متغیرها را استخراج کنند و در نتیجه دقت بالاتری نسبت به سایر مدل‌های مورد بررسی نشان می‌دهند. مدل RF نیز به دلیل استفاده از روش گروهی و کاهش واریانس از طریق bagging عملکرد قابل قبولی دارد، اما نسبت به روش‌های بوستینگ ذکر شده در یادگیری روابط پیچیده ضعیف‌تر عمل می‌کند (Nogueira et al., 2026). اما عملکرد مدل SVR

به انتخاب تابع کرنل و تنظیم پارامترها حساس است و بنابراین در سیستم‌هایی با رفتار بسیار متغیر (مانند نفوذ عمقی در خاک شنی) کاهش پایداری شدیدتری را نشان می‌دهد (Zhang et al., 2025). مدل رگرسیونی Elnet نیز یک مدل خطی با منظم‌سازی است و فرض اصلی آن تقریب خطی بین متغیرهای ورودی و هدف است (Zou & Hastie 2005). درحالی‌که رابطه بین متغیرها در شبیه‌سازی الگوی توزیع رطوبت غیرخطی است و اثرات متقابل بین متغیرهایی مانند حجم آب، زمان و ویژگی‌های خاک قابل توجه است. در نتیجه ضعیف‌ترین نتایج در مدل رگرسیون Elnet مشاهده شد که این امر نشان‌دهنده غالب بودن رفتار غیرخطی سیستم به‌ویژه در خاک Loamy Sand است. عامل دوم تأثیرگذاری رفتار هیدرولیکی خاک در نتایج حاصل است. در خاک‌هایی با درصد شن بیشتر و منافذ درشت‌تر، هدایت هیدرولیکی اشباع بیشتر و جریان عمدتاً تحت تأثیر نیروی گرانش است. نفوذ آب در عمق تحت چنین شرایطی نسبت به تغییرات کوچک در عواملی مانند دبی یا رطوبت اولیه خاک بسیار حساس است و این مطلب منجر به غیرخطی‌تر شدن و ناپایداری‌تر شدن رابطه بین متغیرهای ورودی و عمق نفوذ پیاز رطوبتی می‌شود. در نتیجه دقت مدل‌هایی مانند Elnet در مقایسه با مدل‌های درخت – پایه به شدت کاهش می‌یابد. اما در خاک با درصد رس بیشتر و هدایت هیدرولیکی کمتر، مکش ماتریک بیشتر و رفتار جریان آرام‌تر و قابل پیش‌بینی‌تر است. در نتیجه مدل‌ها بهتر آموزش می‌بینند و دقت در تخمین عمق نفوذ افزایش می‌یابد که نتایج حاصله نیز موید این مطلب است و با نتایج برخی از تحقیقات انجام شده از جمله (Taheri et al., 2025; Alahmad et al., 2025; Priyanka et al., 2024) همخوانی دارد.

ب) عرض سطحی الگوی توزیع رطوبت در خاک

نتایج حاصل از محاسبه میانگین و انحراف معیار شاخص‌های آماری در تکرارهای اعتبارسنجی برای مدل‌های مورد بررسی در بافت Loamy Sand در دو مرحله آزمون و تست جهت تخمین عرض سطحی پیاز رطوبتی در جدول (۶) ارائه شده است. در این بافت بیشترین دقت بین تکرارهای اعتبارسنجی در مرحله تست در مدل XGBoost با مقادیر $R^2=0.97 \pm 0.01$ ، $RMSE = 1/46 \pm 0/45$ cm، $MAE = 0/97 \pm 0/17$ مشاهده شد. کمترین دقت نیز در مدل رگرسیونی Elnet با مقادیر $R^2=0.76 \pm 0.02$ ، $RMSE = 5/07 \pm 0/69$ cm، $MAE = 3/87 \pm 0/29$ cm مشاهده شد. دقیق‌ترین نتایج مدل XGBoost در مرحله تست در تکرار اعتبارسنجی ۲ با مقادیر $R^2=0.98$ ، $MAE = 0/80$ cm، $RMSE = 1/17$ cm مشاهده شد. در مدل رگرسیونی Elnet نیز دقیق‌ترین نتایج در مرحله تست در تکرار اعتبارسنجی ۵ با مقادیر $R^2=0.76$ ، $MAE = 4/16$ cm، $RMSE = 5/84$ cm مشاهده شد.

جدول ۶. شاخص‌های آماری محاسبه شده مدل‌های مذکور در مرحله آموزش و تست در بافت Loamy Sand

ML Models	Width					
	Train (Mean±SD Across Folds)			Test (Mean±SD Across Folds)		
	R ²	RMSE (cm)	MAE (cm)	R ²	RMSE (cm)	MAE (cm)
CatBoost	0.99±0.09	0.20±0.02	0.15±0.01	0.97±0.01	1/68±0/53	1/06±0/20
XGBoost	0.99±0.09	0.08±0.01	0.05±0.00	0.97±0.01	1/46±0/45	0/97±0/17
RF	0.99±0.06	0.94±0.07	0.70±0.05	0.94±0.01	2/43±0/35	1/82±0/20
SVR	0.91±0.01	3/11±0/24	2/00±0/13	0.90±0.03	3/23±0/75	2/30±0/45
Elnet	0.77±0.01	4/99±0/13	3/79±0/11	0.76±0.02	5/07±0/69	3/87±0/29

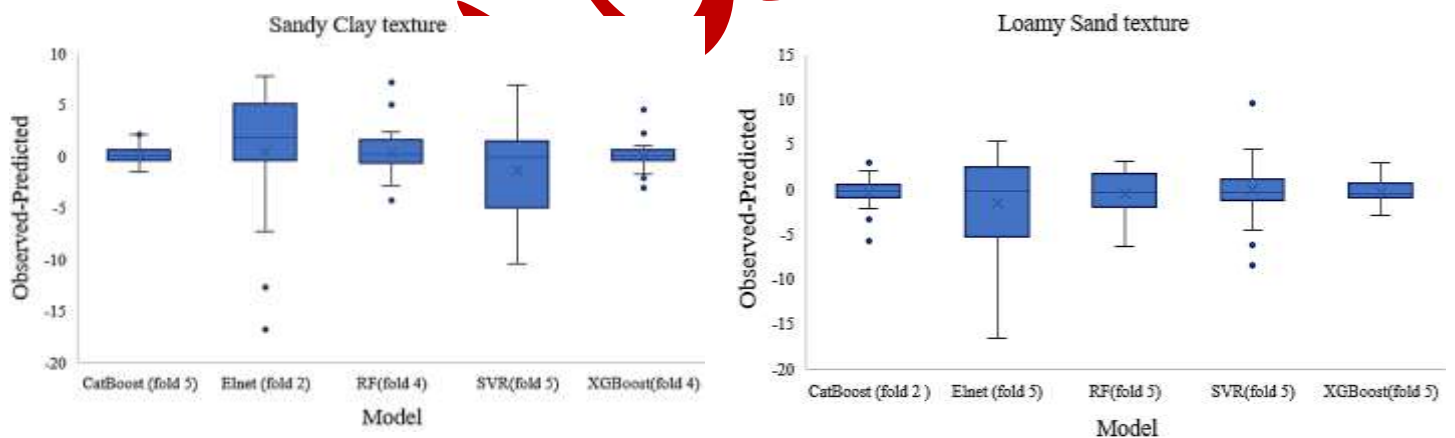
بررسی نتایج ارائه شده در جدول (۷) نیز نشان داد که بیشترین دقت در تخمین عرض سطحی پیاز رطوبتی در بافت Sandy Clay در مرحله تست و در مدل CatBoost با مقادیر $R^2=0.97 \pm 0.03$ ، $RMSE = 1/41 \pm 0/71$ cm، $MAE = 0/91 \pm 0/35$ cm مشاهده شد که با نتایج مدل XGBoost نزدیکی زیادی داشت و کمترین دقت در مدل رگرسیونی Elnet با مقادیر $R^2=0.75 \pm 0.02$ ، $RMSE = 6/07 \pm 0/48$ cm، $MAE = 4/67 \pm 0/26$ cm مشاهده شد. بررسی شاخص‌های آماری محاسبه شده برای هر تکرار اعتبارسنجی نیز نشان داد بهترین

نتایج در مرحله تست در تکرار اعتبارسنجی ۳ مدل CatBoost با مقادیر $R^2=0.98$ ، $RMSE=0.64$ cm، $MAE=0.49$ cm مشاهده شد. در مدل رگرسیونی Elnet نیز بهترین نتایج در تکرار اعتبارسنجی ۲ با مقادیر $R^2=0.77$ ، $RMSE=6.05$ cm، $MAE=4.61$ cm مشاهده شد.

جدول ۷. شاخص‌های آماری محاسبه شده مدل‌های مذکور در مرحله آموزش و تست در بافت Sandy Clay

ML Models	Width					
	Train (Mean±SD Across Folds)			Test (Mean±SD Across Folds)		
	R ²	RMSE (cm)	MAE (cm)	R ²	RMSE (cm)	MAE (cm)
CatBoost	0.99±0.05	0.17±0.02	0.13±0.01	0.97±0.03	1.41±0.71	0.91±0.35
XGBoost	0.98±0.06	0.62±0.00	0.45±0.00	0.97±0.01	1.63±0.32	1.07±0.20
RF	0.97±0.03	1.13±0.11	0.74±0.04	0.94±0.02	2.96±0.40	1.96±0.20
SVR	0.87±0.01	4.41±0.17	2.95±0.02	0.85±0.01	4.69±0.26	2.20±0.11
Elnet	0.76±0.00	6.03±0.10	4.62±0.10	0.75±0.02	6.07±0.48	4.67±0.26

مقایسه نمودار جعبه‌ای خطای پیش‌بینی نیز نشان داد در بافت Loamy Sand هر دو مدل CatBoost و XGBoost عملکرد نسبتاً مشابهی داشتند. با این حال در مدل XGBoost پراکندگی خطای کمتر و RMSE پایین‌تر و نبود نقاط پرت مشاهده شد که بیانگر دقت و پایداری بالاتر آن نسبت به مدل CatBoost است. در بافت Sandy Clay نیز بیشترین دقت در مدل CatBoost مشاهده شد و ضعیف‌ترین عملکرد در هر دو بافت در مدل Elnet مشاهده شد (شکل ۵). همچنین بررسی شاخص‌های آماری حاصل از اعمال OOF بر عرض سطحی پیاز رطوبتی نشان داد که مدل CatBoost در هر دو بافت دقیق‌ترین نتایج را ارائه کرده است که مشابهت زیادی با نتایج مدل XGBoost دارد و ضعیف‌ترین نتایج در مدل Elnet مشاهده شد (جدول ۸). در خاک Loamy Sand با درصد شن بالاتر تغییرات عرض سطحی به دلیل کاهش مکش ماتریک نسبت به عمق کمتر است و تابع مستقیم‌تری از متغیرهایی مانند دبی و زمان است. در نتیجه رابطه بین متغیرهای ورودی و عرض سطحی رفتار ساده‌تر و منظم‌تری دارد و مدل‌ها بهتر آموزش می‌بینند.

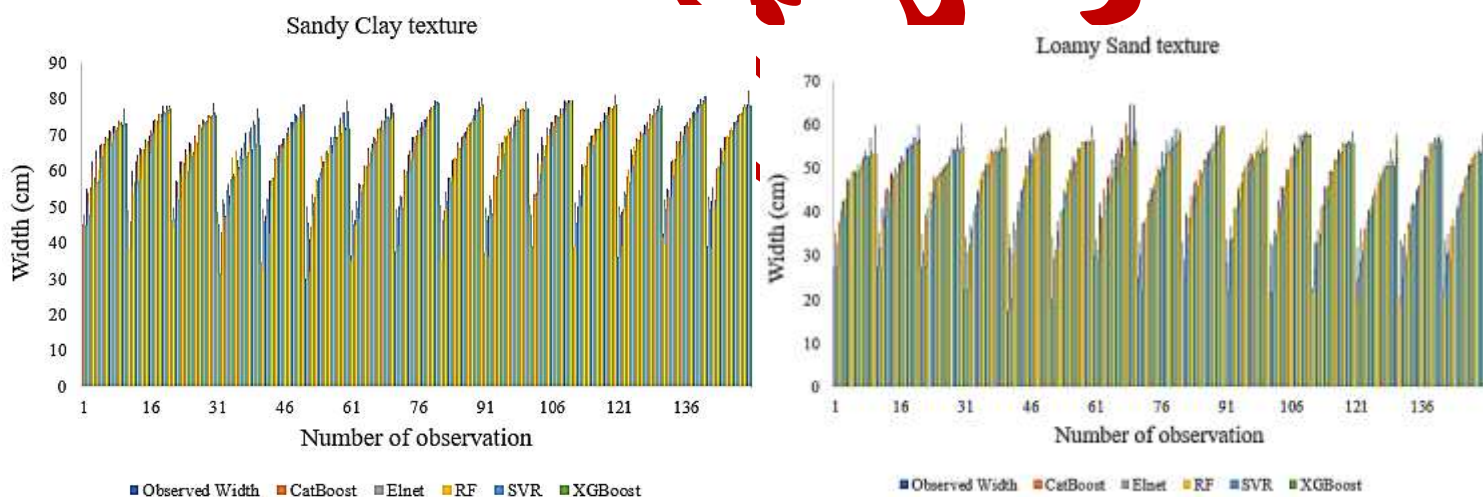


شکل ۵. نمودار جعبه‌ای خطای پیش‌بینی مدل‌های مورد بررسی در تخمین عرض سطحی پیاز رطوبتی

جدول ۸. عملکرد مدل‌های مورد بررسی در تخمین عرض سطحی پیاز رطوبتی بر اساس پیش‌بینی‌های خارج از تکرار اعتبارسنجی (OOE)

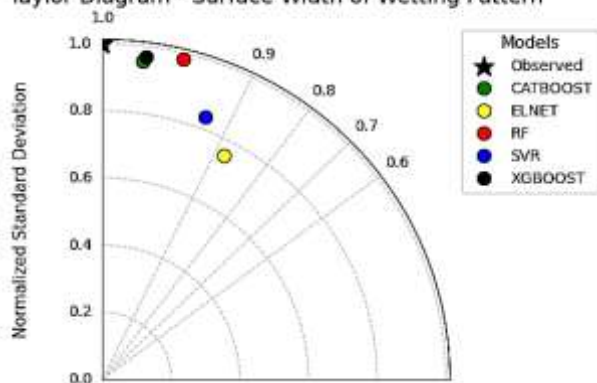
بافت خاک ML	Sandy Clay			Loamy Sand		
	R ²	RMSE (cm)	MAE (cm)	R ²	RMSE (cm)	MAE (cm)
CatBoost	۰/۹۸	۱/۵۸	۰/۹۱	۰/۹۷	۱/۷۶	۱/۰۶
XGBoost	۰/۹۸	۱/۶۷	۱/۰۷	۰/۹۷	۱/۵۳	۰/۹۷
RF	۰/۹۴	۲/۹۸	۱/۹۶	۰/۹۴	۲/۴۵	۱/۸۲
SVR	۰/۸۵	۴/۷۱	۳/۳۰	۰/۹۰	۳/۳	۲/۳۰
Elnet	۰/۷۶	۶/۰۹	۴/۶۷	۰/۷۶	۵/۱۲	۳/۸۷

نمودارهای مقایسه عملکرد مدل‌های مورد بررسی و تیلور شاخص‌های محاسبه شده نیز دقت مدل CatBoost را در تخمین عرض سطحی پیاز رطوبتی تایید می‌کند (شکل ۶-۷). همچنین نشان می‌دهد که مدل‌های SVR و Elnet از دقت کمتری در مقایسه با مدل‌های درخت-پایه در تخمین عرض سطحی در هر دو بافت برخوردارند و ضعیف‌ترین نتایج در مدل رگرسیونی Elnet و در بافت Sandy Clay مشاهده شد.



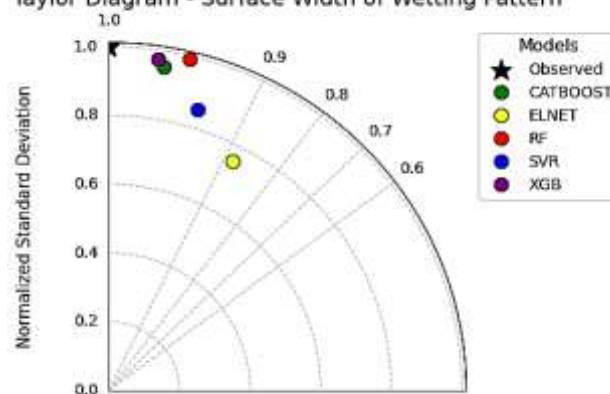
شکل ۶. مقایسه عملکرد مدل‌های مورد بررسی در پیش‌بینی عرض سطحی پیاز رطوبتی

Taylor Diagram - Surface Width of Wetting Pattern



Sandy Clay texture

Taylor Diagram - Surface Width of Wetting Pattern



Loamy Sand texture

شکل ۷. نمودار تیلور عملکرد مدل‌های مورد بررسی در تخمین عرض سطحی پیاز رطوبتی

در مجموع در خاک Sandy Clay حرکت آب در هر دو جهت کنترل شده‌تر است و تغییرات سیستم نرم‌تر و پیوسته‌تر رخ می‌دهد. در نتیجه رابطه بین متغیرهای ورودی و هدف ساختارمندتر است و عملکرد مدل در هر دو صفت باثبات‌تر است (Rajhi et al., 2026) که در سایر تحقیقات از جمله (Mirzaeitalarposhti et al., 2022; Kheimi et al., 2025) نیز تایید شده است. بررسی نتایج حاصل از سایر تحقیقات انجام شده در زمینه تخمین ابعاد پیاز رطوبتی نیز تایید کننده دقت مدل‌های غیرخطی در شرایط مختلف است که با نتایج تحقیق حاضر همخوانی دارد. از جمله (Shiri et al., 2020) دقت مدل‌های جنگل تصادفی و برنامه‌ریزی بیان ژن را در بافت‌های مختلف و تحت سیستم آبیاری قطره‌ای سطحی و زیرسطحی تایید کردند. (Karimi et al., 2020) نیز دقت مدل ANN و رگرسیون غیرخطی را در مقایسه با یک مدل تجربی ایجاد شده بر مبنای آنالیز ابعادی در برآورد مساحت خیس شده اطراف قطره‌چکان تایید کردند. همچنین (Kisi et al., 2021) دقت مدل هیبریدی ANFIS-FCM-PSOGSA را در تخمین ابعاد پیاز رطوبتی تحت شرایط آبیاری پیوسته و پالسی در آبیاری قطره‌ای تایید کردند.

نتیجه‌گیری و پیشنهادها

در تحقیق حاضر به ارزیابی دقت ۵ مدل یادگیری ماشین در تخمین عرض سطحی و عمق نفوذ پیاز رطوبتی در دو بافت مختلف خاک و در شرایط کاربرد آب شور پرداخته شد و از شاخص‌های آماری R^2 , RMSE, MAE جهت بررسی عملکرد مدل‌ها استفاده شد. نتایج بدست آمده نشان داد که در مجموع مدل‌های درخت-پایه از دقت قابل قبولی در تخمین عمق نفوذ و عرض سطحی پیاز رطوبتی در هر دو بافت برخوردارند. هر چند آموزش مدل‌ها به واسطه رفتار غیرخطی‌تر نفوذ در خاک سنی، برای تخمین صفت عمق نفوذ ناپایدارتر بوده و نتایج حاصله این مطلب را تایید می‌کند. دقیق‌ترین تخمین عمق نفوذ ($R^2=0.94$, $RMSE=0.27$ cm, $MAE=0.10$ cm) و عرض سطحی پیاز رطوبتی ($R^2=0.98$, $RMSE=1.58$ cm, $MAE=0.91$ cm) در مدل CatBoost و در بافت Sandy Clay مشاهده شد. ضعیف‌ترین نتایج برای صفات مذکور نیز در هر دو بافت در مدل Elnet مشاهده شد. از محدودیت‌های این تحقیق می‌توان به حجم محدود داده‌های در دسترس اشاره کرد که به واسطه گردآوری داده‌ها در شرایط آزمایشگاهی بوده و پیشنهاد می‌شود در تحقیقات آینده از مجموعه وسیع‌تری از اطلاعات در بافت‌های مختلف استفاده گردد و به‌علاوه به منظور شبیه‌سازی شرایط واقعی، رطوبت خاک با استفاده از سنسورهای دائمی ثبت و به عنوان متغیر ورودی مورد استفاده قرار گیرد. در این تحقیق برای کنترل اثرات شوری تجمعی در خاک از ایشوئی استفاده شد، اما همچنان احتمال وجود نمک و اثرگذاری بر ابعاد پیاز رطوبتی وجود دارد. به علاوه قرار رفتن قطره‌چکان در نزدیکی صفحه شفاف باعث گسترش نامتقارن پیاز رطوبتی می‌شود که پیشنهاد می‌شود در تحقیقات آینده مورد مطالعه قرار گیرد.

منابع

- سعیدی عباس آباد، میلاد؛ گلستانی کرمانی، سودابه؛ مهبیجی نصر آبادی، مهدی؛ ذونعمت کرمانی، محمد. (۱۴۰۳). تاثیر میدان مغناطیسی بر الگوی توزیع آب شور تحت زمان در آبیاری قطره‌ای. *نشریه آبیاری و زهکشی ایران*، ۱ (۱۸)، ۲۰۴-۱۸۵.
- کاندلوس، مازیار؛ لیاقت، عبدالمجید؛ عباسی، فریدون. (۱۳۸۷). برآورد الگوی رطوبتی خاک در آبیاری قطره‌ای زیرسطحی با استفاده از روش تحلیل ابعادی. *مجله علوم کشاورزی ایران*، (۲)، ۳۹، ۳۷۸-۳۷۱.
- نیکبخت، جعفر؛ عبدالمجید سیاهکلودی، مرضیه. (۱۳۹۳). اثر مغناطیسی کردن آب آبیاری بر ویژگی‌های الگوی رطوبتی در آبیاری قطره‌ای سطحی. *نشریه دانش آب و خاک*، (۴)، ۱۵۲، ۲۴-۱۳۹.

REFERENCES

- Alahmad, T., Neményi, M., Széles, A., Ali, NA., Hijazi, O., & Nyéki, A. (2025). Spatiotemporal prediction of soil moisture content at various depths in three soil types using machine learning algorithms. *Frontiers in Soil Science*, 5, 1612908. <https://doi.org/10.3389/fsoil.2025.1612908>.
- Al-Ogaidi, AAM., Wayayok, A., Kamal, MR., & Abdullah, AF. (2015). A modified empirical model for estimating the wetted zone dimensions under drip irrigation. *Journal Teknologi*, 76, 69–73.
- Al-Ogaidi, AAM., Wayayok, A., Rowshon, MK., & Abdullah, AF. (2016). Wetting patterns estimation under drip irrigation systems using an enhanced empirical model. *Agricultural Water Management*, 176, 203–213.
- Amin, MSM., & Ekhmaj, AIM. (2006). DIPAC-Drip irrigation water distribution pattern calculator. 7th International micro irrigation congress, 10–16 September, PWTC, Kuala Lumpur, Malaysia.
- Arbat, G., Puig-Bargués, J., Duran-Ros, M., Barragán, J., & Ramírez de Cartagena, F. (2013). Drip-Irrigation: computer software to simulate soil wetting patterns under surface drip irrigation. *Computers and Electronics in Agriculture*, 98, 183–192.
- Bentéjac, C., Csörgő, A., & Martínez-Muñoz, G. (2021). A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review*, 54, 1937–1967.
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794.
- Cook, FJ., Thorburn, PJ., Fitch, P., & Bristow, KL. (2003). WetUp: a software tool to display approximate wetting patterns from drippers. *Irrigation Science*, 22, 129-134.
- Cristóbal-Muñoz, I., Prado-Hernández, JV., Martínez-Ruiz, A., Pascual-Ramírez, F., Cristóbal-Acevedo, D., & Cristóbal-Muñoz, D. (2022). An improved empirical model for estimating the geometry of the soil wetting front with surface drip irrigation. *Water*, 14(11), 1827. <https://doi.org/10.3390/w14111827>.
- Delgado, JA., Short, NM., Roberts, DP., & Vandenberg, B. (2019). Big data analysis for sustainable agriculture on a geospatial cloud framework. *Frontiers in Sustainable Food Systems*, 3, Article 469303. <https://doi.org/10.3389/fsufs.2019.00054>

- Dhaliwal, JK., Panday, D., Saha, D., Lee, J., Jagadamma, S., Schaeffer, S., & Mengistu, A. (2022). Predicting and interpreting cotton yield and its determinants under long-term conservation management practices using machine learning. *Computers and Electronics in Agriculture*, 199, doi:[10.1016/j.compag.2022.107107](https://doi.org/10.1016/j.compag.2022.107107).
- Elmaloglou, S., Soulis, KX., & Dercas, N. (2013). Simulation of soil water dynamics under surface drip irrigation from equidistant line sources. *Water Resource Management*, 27, 4131-4148.
- Hammami, M., & Zayani, K. (2016). An analytical approach to predict the moistened bulb volume beneath a surface point source. *Agricultural Water Management*, 166, 123-129.
- Hanson, B.R. Gratten, S.R. & Fulton, A. (2006). Agricultural salinity and drainage. Regents of the University of California, Oakland, 180.
- Hong, H., Pourghasemi, HR., & Pourtaghi, ZS. (2016). Landslide susceptibility assessment in lianhua county (China): a comparison between a random forest data mining technique and bivariate and multivariate statistical models, *Geomorphology*, 259, 105–118.
- Kandelous, M., Liaghat, A., & Abbasi, F. (2008). Estimation of soil moisture pattern in subsurface drip irrigation using dimensional analysis methods. *The Journal of Agricultural Science*, 39(2), 371-378. (In Persian).
- Kandelous, MM., & Šimůnek, J. (2010). Comparison of numerical, analytical, and empirical models to estimate wetting patterns for surface and subsurface drip irrigation. *Irrigation Science*, 28, 435–444.
- Karimi, B., Mohammadi, P., Sanikhani, H., Salih, SQ., & Yaseen, ZM. (2020). Modeling wetted areas of moisture bulb for drip irrigation systems: An enhanced empirical model and artificial neural network. *Computers and Electronics in Agriculture*, 178, 105767.
- Kheimi, M., Alotaibi, F., & Alqahtani, A. (2025). Conventional and advanced ai-based models in soil moisture prediction. *Journal of Hydrology*, (In press). <https://doi.org/10.1016/j.jhydrol.2025.XXXXXX>.
- Kisi, O., Khosravinia, P., Heddam, S., Karimi, B., & Karimi, N. (2021). Modeling wetting front redistribution of drip irrigation systems using a new machine learning method: adaptive neuro-fuzzy system improved by hybrid particle swarm optimization-gravity search algorithm. *Agricultural Water Management*, 256. 107067.
- Kusumavathi, K., Konatala, R., Lai, P., Sarkar, S., Banerjee, H., Bandopadhyay, P., Sethi, D., & Upendar, K. (2025). Artificial intelligence for fostering sustainable agriculture. *Gurrent Plant Biology*, 42, 100476.
- Malek, K., & Peters, RT. (2011). Wetting pattern models for drip irrigation, new empirical models. *Journal of Irrigation and Drainage Engineering*, 137, 530-536.
- Mirzaeitalarposhti, R., Shafizadeh-Moghadam, H., & Demyan, MS. (2022). Digital soil texture mapping and spatial transferability of machine learning models using sentinel-1, sentinel-2, and terrain-derived covariates. *Remote Sensing*, 14(23), 5909. <https://doi.org/10.3390/rs14235909>.
- Moncef, H., & Khemaies, Z. (2016). An analytical approach to predict the moistened bulb volume beneath a surface point source. *Agricultural Water Management*, 166, 123–129.

- Nikbakht, J., & Abdollahi Siahkalroudi, M. (2014). Effect of magnetization of irrigation water on the properties of soil wetting pattern in surface drip irrigation. *Water and Soil Science*, 24(4), 139-152. (In Persian).
- Nogueira, LSR., De Carvalho, MAS., Santos, BDO., Yonaba, R., Bamal, A., Uddin, MG., Bodini, M., & Goliatt, L. (2026). A comparative study of ensemble and non-ensemble machine learning methods for predicting river pollution index. *Ecological Informatics*, 81, 103617. <https://doi.org/10.1016/j.ecoinf.2025.103617>.
- Priyanka, P., Kumar, P., & Panda, S. (2024). Can machine learning models predict soil moisture evaporation rates? an investigation via novel feature selection techniques and model comparisons. *Frontiers in Earth Science*, 12, 1344690. <https://doi.org/10.3389/feart.2024.1344690>.
- Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, AV., & Gulin, A. (2018). Catboost: unbiased boosting with categorical features. Proceedings of the 32nd International Conference on Neural Information Processing Systems, Montréal, 3-8 December, 6639-6649.
- Rajhi, M., Deak, T., & Dobos, E. (2026). Non-invasive soil texture prediction using machine learning and multi-source environmental data. *Soil Systems*, 10(1), 8. <https://doi.org/10.3390/soilsystems10010008>
- Saeidi Abbasabad, M., Golestani Kermani, S., Mohayeji Nasrabadi, M., & Zounemat-Kermani, M. (2024). Effect of magnetic fields on saline water distribution pattern under time in drip irrigation. *Iranian Journal of Irrigation and Drainage*, 1(18), 185-204. (In Persian).
- Samadianfard, S., Sadraddini, AA., Nazemi, AH., Provenzano, G., & Kisi, O. (2012). Estimation soil wetting pattern for drip irrigation using genetic programming. *Spanish Journal of Agricultural Research*, 10(4), 1155-1166.
- Segovia, JA., Toaquiza, JF., Llanos, JR., & Rivas, DR. (2023). Meteorological variables forecasting system using machine learning and open-source software. *Electronics*, 12(4), 1007.
- Seifu Majdar, R., Rahnamaei, A., & Babazadeh, V. (2025). Hybrid machine learning in hydrological runoff forecasting: an exploration of extreme gradient-boosting and categorical gradient boosting optimization in the russian river basin. *Advances in Engineering and Intelligence Systems*, 4(2). <https://doi.org/10.22034/aeis.2025.509199.1293>.
- Sejna, M., Simunek, J., & Van Genuchten, MT. (2014). The HYDRUS software package for simulating two and three dimensional movement of water, heat and multiple solutes in variably – saturated porous media, version 2-04. (PC Progress, Prague, Czech Republic).
- Shiri, J., Karimi, B., Karimi, N., Kazemi, MH., & Karimi, S. (2020). Simulating wetting front dimensions of drip irrigation systems: multi criteria assessment of soft computing models. *Journal of Hydrology*, 585, 124792.
- Sishodia, RP., Ray, RL., & Singh, SK. (2020). Applications of remote sensing in precision agriculture: a review. *Remote Sensing*, 12(19), 1–31.
- Smola, AJ., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, 14(3), 199–222.

- Taheri, M., Bigdeli, M., Imanian, H., & Mohammadiam, A. (2025). An overview of machine-learning methods for soil moisture estimation. *Water*, 17(11), 1638. <https://doi.org/10.3390/w17111638>
- Vanwinckelen, G., & Blockeel, H. (2012). On estimating model accuracy with repeated cross-validation. *Proceedings of the 21st Belgian-Dutch conference on machine learning*, 39-44.
- Vapnik, V. (1984). *Estimation of dependences based on empirical data*. Springer-Verlag, 400 p. <https://books.google.nl/books?id=wxFS0AEACAAJ>
- Vapnik, V., & Chervonenkis, A. (1974). *Theory of pattern recognition*. Nauka, Moscow. 353 p.
- Wang, X., Liu, T., Zheng, X., Peng, H., Xin, J., & Zhang, B. (2018). Short-term prediction of groundwater level using improved random forest regression with a combination of random features. *Applied Water Science*, 8(5), 1–12.
- Zhang, X., Sun, X., & Lin, Z. (2025). Improving soil moisture prediction using gaussian process regression. *Smart Agricultural Technology*, 11, 100905. <https://doi.org/10.1016/j.atech.2025.100905>.
- Zhu, Z., Waseem Rasheed, M., Safdar, M., Yao, B., Tumaerbai, H., Sarwar, A., & Zhu, L. (2024). Intermittent drip irrigation soil wet front prediction model and effective water storage analysis. *Sustainability*, 16, 9553. <https://doi.org/10.3390/su16219553>
- Zou, H., & Hastie, T. (2005). Regularization and variable selection via the Elastic net. *Journal of the Royal Statistical Society Series A*, 67 (2), 301-320.