

# ارزیابی عملکرد مدل‌های مبتنی بر یادگیری ماشین در برآورد حداکثر عمق آبشستگی اطراف دماغه آبشکن نوع باندا لایک

## چکیده

در تحقیق حاضر عملکرد روش‌های مبتنی بر یادگیری ماشین به منظور پیش‌بینی حداکثر عمق آبشستگی اطراف آبشکن نوع باندا لایک مورد ارزیابی قرار گرفت. برای این منظور سه روش مدل جنگل‌های تصادفی (RF)، ماشین بردار پشتیبان (SVM) و روش برنامه‌ریزی بیان ژن (GEP) مورد استفاده قرار گرفت. به منظور آموزش و آزمایش مدل‌ها از ۱۰۸ سری اطلاعات (۸۷ سری برای آموزش و ۲۱ سری برای تست) مستخرج از نتایج یک تحقیق آزمایشگاهی استفاده شد. مدل‌ها با ترکیب‌های متفاوتی (تک متغیره، دو متغیره، سه متغیره و چهار متغیره) از ورودی‌ها (Fr: عدد فرود جریان، S/L: نسبت فاصله به طول آبشکن،  $\theta$ : زاویه نصب آبشکن نسبت به ساحل و  $\alpha$ : تخلخل قسمت نفوذپذیر سازه) مورد ارزیابی قرار گرفتند. نتایج حاصل نشان داد برای تمامی روش‌ها در حالت ورودی تک متغیره، بیشترین و کمترین تاثیر به ترتیب مربوط به پارامترهای  $\alpha$  و S/L بودند. در مدل SVM با افزایش تعداد ورودی‌ها از تک متغیره به دو متغیره میانگین شاخص MAE تقریباً ۲ برابر افزایش یافت. در مدل GEP نیز افزایش تعداد ورودی‌ها از سه متغیره به ۴ متغیره میانگین شاخص MAE تقریباً ۳/۵ برابر افزایش یافت. ولی در روش RF افزایش تعداد ورودی‌ها منجر به افزایش دقت مدل شد و متوسط شاخص MAE در حالت ۴ متغیره نسبت به سه متغیره ۸۳ درصد کاهش یافت. در نهایت مشخص شد روش RF در برآورد عمق آبشستگی اطراف آبشکن نوع باندا لایک از عملکرد بسیار بهتری (RMSE=۰/۰۰۶ و MAE=۰/۰۰۹) نسبت به سایر روش‌ها برخوردار بوده و این مدل با ورودی‌های یکسان از پراکنش خطای کمتری برخوردار بود.

واژه‌های کلیدی: آبشستگی، هوش مصنوعی، رودخانه، آبشکن.

## Evaluation of the performance of machine learning methods for estimating the maximum scour depth around the bandallike spur-dike

### ABSTRACT

In this study, the performance of machine learning-based methods for predicting the maximum scour depth around a Bandallike spur-dike is evaluated. For this purpose, three methods of Random Forest (RF) model, Support Vector Machine (SVM), and Gene Expression Programming (GEP) were used. To train and test the models, 108 data series (87 series for training and 21 series for testing) were extracted from the results of an experimental study. The models were evaluated with four different combinations of inputs (Fr: flow Froude number, S/L: ratio of distance to breakwater length,  $\theta$ : spur-dike installation angle relative to the bank, and  $\alpha$ : porosity of the permeable structure). The results showed that for all methods in the one input mode, the parameters with the most and least impact were, in order,  $\alpha$  and S/L. In the SVM model, the average MAE index increased by about 2 times when the number of inputs increased from one input mode. In the GEP model, the average MAE index increased by about 3.5 times when the number of inputs increased from three to four inputs mode. However, in the RF method, increasing the number of inputs led to an increase in model accuracy, and the average MAE index decreased by 83% in the four inputs mode compared to the three inputs mode. Finally, it was found that the RF method had much better performance (MAE = 0.006 and RMSE = 0.009) in estimating the scour depth around the Bandal-like spur-dike than the other methods, and this model had less error spread with the same inputs.

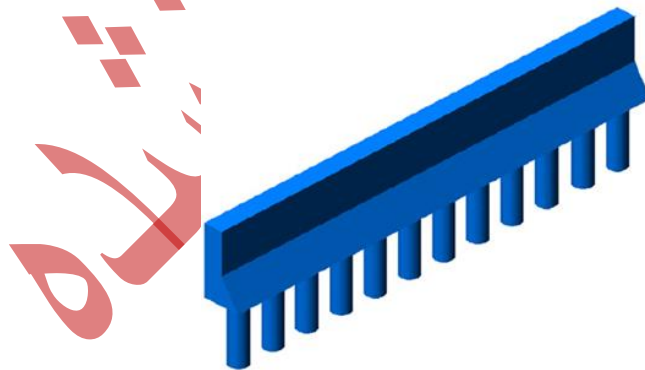
**Keywords:** Scouring, Artificial intelligence, River, spur-dike.

## مقدمه

مسیرهای طبیعی رودخانه به عنوان مجرای طبیعی جمع‌آوری آب‌های سطحی و قسمتی از آب‌های زیرزمینی می‌باشد. شکل‌گیری جوامع بشری همواره در مجاورت رودخانه‌ها بوده و امروزه نیز جوامع متعددی در ساحل رودخانه‌ها قرار دارند. این موضوع باعث گسترش صنایع متفاوتی در اطراف رودخانه‌ها در سراسر جهان شده است. ایجاد تاسیسات در کنار رودخانه‌ها همواره مورد

تهدید خطرانی بوده که می تواند خسارتها و صدمات جبران ناپذیری را به انسان ها وسازه های اطراف خود وارد نماید. از جمله مهمترین این خطرها وقوع سیلاب می باشد. یکی دیگر از خطرات بالقوه در این راستا فرسایش سواحل و در نتیجه ناپایداری اراضی اطراف رودخانه ها می باشد. بدین صورت که بستر رودخانه با گذشت زمان بر اثر فرسایش سواحل به اطراف پیشروی نموده و برای اراضی اطراف ایجاد مشکل می نماید. لذا مطالعه در خصوص حفاظت از اراضی اطراف رودخانه و کنترل فرسایش از گذشته مورد مطالعه بسیاری از محققین قرار گرفته است. در این راستا روش های مختلفی در خصوص کنترل فرسایش سواحل رودخانه ها ارائه شد. یکی از مرسوم ترین روش ها در این خصوص استفاده از سازه آبشکن می باشد. آبشکن ها با انتقال الگوی جریان فرسایش پذیر از سواحل و کناره ها و متمایل کردن آن به خط الرأس رودخانه زمینه مناسبی جهت رسوب گذاری و رسوبگذاری در بین آبشکن ها را فراهم می کند (روشنگر و همکاران، ۱۴۰۲). تا کنون مطالعات متعددی در خصوص هیدرولیک جریان و رسوب اطراف آبشکن ها انجام شده است (Pandey et al., 2017).

یکی از مسائل اساسی در خصوص سازه آبشکن، آبستتگی اطراف دماغه آن به واسطه تغییر ایجاد شده در الگوی جریان می باشد. وقوع آبستتگی اطراف دماغه منجر به ایجاد ناپایداری سازه و در نهایت تخریب آن می شود. لذا ارائه راهکارهایی جهت کنترل و کاهش آبستتگی اطراف دماغه آبشکن از اهمیت بالایی برخوردار بوده و تا کنون روش های اجرایی متعددی مانند انجام ریپ رپ و همچنین استفاده از آبشکن محافظ بدین منظور ارائه شده است. هنگامی که جریان به بالادست آبشکن نفوذناپذیر برخورد می کند یک جریان رو به پایین ایجاد می شود که با توجه به جهت جریان به سمت دماغه آبشکن متمایل شده و منجر به ایجاد آبستتگی در دماغه می شود. همچنین با برخورد جریان به دماغه آبشکن پدیده جدایی جریان رخ داده که منجر به ایجاد ناحیه کم فشار و جابه جایی رسوبات در این ناحیه خواهد شد. به منظور کاهش عمق آبستتگی اطراف دماغه آبشکن، سازه آبشکن جدیدی به نام باندا لایک معرفی شد. این سازه برای اولین بار در کشورهای جنوب و جنوب شرقی آسیا به منظور ثابت نگه داشتن عمق جریان برای کشتیرانی مورد استفاده قرار گرفت و با انتقال جریان های مخرب به مرکز رودخانه از فرسایش کناره ها و سواحل به صورت غیر مستقیم محافظت به عمل می آورد (Teraguchi et al., 2011). ساختار این سازه بدین صورت بود که از یک آبشکن نفوذ پذیر در قسمت پایین و یک آبشکن نفوذ ناپذیر روی آن ساخته شده است. در نتیجه با توجه به الگوی جریان اطراف آبشکن نفوذناپذیر میزان جریان های پایین رونده کاهش یافته و لذا عمق حداکثر آبستتگی اطراف دماغه کاهش می یابد (شکل ۱).



شکل ۱. شکل شماتیک سازه باندا لایک

در زمینه آبستتگی اطراف دماغه آبشکن تا کنون مطالعات متعددی به صورت آزمایشگاهی و عددی توسط محققین مختلف انجام شده است (Pandey et al., 2017). با توجه به پیچیده بودن مکانیسم حرکت رسوبات در پدیده آبستتگی و همچنین تأثیر داشتن پارامتر های زیاد در این پدیده امکان ارائه روابط تجربی دقیق در این زمینه وجود ندارد. همچنین شبیه سازی این پدیده در آزمایشگاه بسیار پرهزینه و زمان بر می باشد. روابط تجربی ارائه شده در این راستا دارای محدودیت هایی بوده و در تمامی شرایط دارای دقت قابل قبول نمی باشند. لذا در چند سال گذشته استفاده از روش های یادگیری ماشین در خصوص تخمین آبستتگی رسوبات اطراف سازه های گوناگون احداث شده در آبراهه ها از جمله آبشکن توسعه پیدا کرده است.

فرزین و همکاران (۱۳۹۸) تخمین آبستتگی اطراف سازه آبشکن را با استفاده از روش های هوش مصنوعی مورد مطالعه قرار

دادند. روش‌های مورد استفاده در این تحقیق برنامه‌ریزی بیان ژن (GEP) و روش گروهی مدلسازی داده‌ها (GMDH) بود. نتایج حاصل نشان داد که در بخش آموزش و صحت‌سنجی مدل GMDH نسبت به مدل GEP دارای دقت بالاتری می‌باشد. (Basser et al., 2014) با استفاده از روش‌های یادگیری ماشین به بررسی اهمیت و تعیین پارامترهای بهینه سازه آبشکن حافظتی روی کنترل آبشستگی اطراف دماغه آبشکن‌های اصلی پرداختند. این پارامترها شامل زاویه نصب، طول و فاصله آبشکن حافظتی از آبشکن اصلی، شدت جریان و اندازه رسوبات می‌باشد. برای این منظور از روش‌های SVR، ANFIS و ANN استفاده شد. نتایج حاصل نشان داد که روش SVR عملکرد بهتری در این زمینه نسبت به دو روش دیگر دارا می‌باشد. (Pandey et al., 2015) به بررسی دقت روابط موجود در پیش‌بینی حداکثر عمق آبشستگی اطراف سازه آبشکن صلب پرداختند. آنها از اطلاعات و نتایج سایر محققین در تحقیقات پیشین استفاده نموده و با استفاده از روش‌های آماری معادلات جدیدی را در این خصوص توسعه دادند.

Parsaie et al., (2019) در تحقیقی به تخمین عمق آبشستگی در زیر خطوط لوله در رودخانه‌ها را با استفاده از ماشین بردار پشتیبان (SVM) پرداختند. در این تحقیق از داده‌های میدانی استفاده شد. همچنین عملکرد مدل SVM با استفاده از شاخص‌های آماری با مدل‌های شبکه عصبی مصنوعی (ANN) و سیستم‌های استنتاج فازی عصبی تطبیقی (ANFIS) مقایسه شد. نتایج حاصل نشان از برتری مدل SVM نسبت به دو مدل دیگر به منظور تخمین عمق آبشستگی در زیر خطوط لوله در رودخانه‌ها داشت. همچنین انجام آنالیز حساسیت نشان داد که زبری نسبی، عدد فرود و تنش برشی مهمترین برای یافتن موثرترین پارامترها در این تخمین می‌باشند.

Najafzadeh and Oliveto (2021) در تحقیقی به بررسی آبشستگی اطراف گروه‌های شمع تحت جریان‌های ثابت با استفاده از مدل‌های هوش مصنوعی (AI) پرداختند. در این تحقیق مدل‌های GEP، EPR، MARS و M5MT برای توسعه معادلات رگرسیون غیرخطی به منظور تخمین حداکثر عمق آبشستگی تعادلی مورد استفاده قرار گرفتند. نتایج حاصل نشان داد که تکنیک MARS تخمین‌های دقیق‌تری از عمق آبشستگی ارائه می‌دهد.

Nou et al., (2022) دقت تخمین آبشستگی اطراف گروه دو پایه و گروه سه پایه مربوط به پل را با استفاده از روش یادگیری ماشین مورد بررسی قرار دادند. در این مطالعه از روش ORELM استفاده شد. نتایج حاصل نشان داد که این مدل دقت قابل قبولی در این زمینه دارد. همچنین میزان تاثیرگذاری پارامترهای تاثیرگذار نیز مورد ارزیابی قرار گرفت. (Pandey et al., 2022) روی پیش‌بینی تغییرات زمانی عمق آبشستگی اطراف سازه آبشکن در حضور رسوبات چسبنده با استفاده از روش‌های یادگیری ماشین مطالعه کردند. این روش‌ها شامل درخت تصمیم‌گیری (GBDT)، شبکه عصبی پیشخور (CFNN) و روش رگرسیون کرنل (KRR) بود. نتایج حاصل نشان داد که روش GBDT از دو روش دیگر با دقت بالاتری قادر به تخمین تغییرات زمانی عمق آبشستگی می‌باشد.

Tripathi and Pandey (2023) با استفاده از مدل GEP حداکثر عمق آبشستگی اطراف آبشکن نفوذناپذیر را پیش‌بینی نموده و دقت بالای این مدل در این خصوص را گزارش دادند.

Kumar et al., (2023) به بررسی دقت روش‌های یادگیری ماشین در تخمین تغییرات زمانی چاله آبشستگی اطراف پایه‌های پل با سطح مقطع دایره‌ای شکل پرداختند. در این تحقیق از دو روش ترکیبی BR و ABR و همچنین از روش SVR استفاده شد. نتایج حاصل نشان داد که روش‌های ترکیبی با دقتی به مراتب بالاتر نسبت به روش SVR قادر به تخمین تغییرات زمانی چاله آبشستگی اطراف پایه پل می‌باشد.

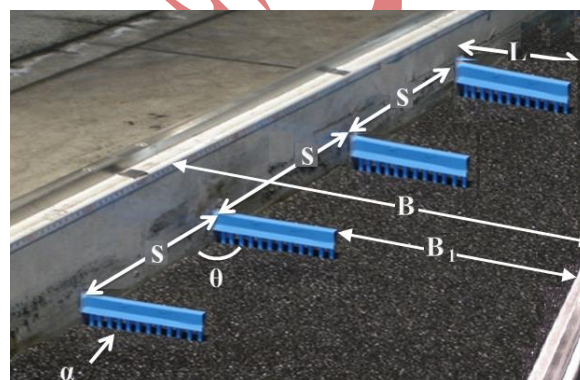
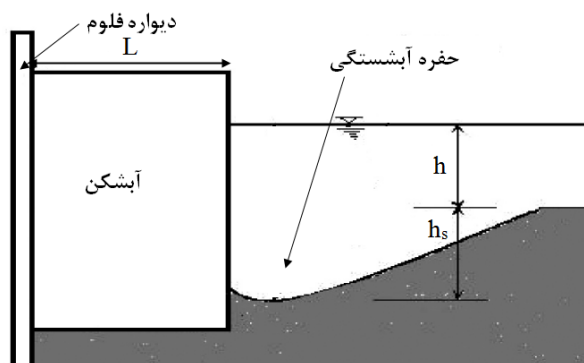
تا کنون روابط تجربی متعددی جهت تخمین حداکثر عمق آبشستگی اطراف دماغه سازه آبشکن نفوذناپذیر به منظور در نظر گرفتن تمهیداتی در راستای طراحی ارائه شده است. ولی در خصوص آبشکن نفوذپذیر روابط زیادی ارائه نشده است. باید توجه داشت روابط تجربی دارای محدودیت کاربرد بوده و در تمامی حالات قادر به تخمین دقیق نمی‌باشند. با توجه به این موضوع که روش‌های مبتنی بر یادگیری ماشین تابع شرایط آزمایشگاه نمی‌باشند در نتیجه استفاده از این روش‌ها در این زمینه می‌تواند سودمند باشد. لذا هدف از تحقیق حاضر ارزیابی عملکرد روش‌های مبتنی بر یادگیری ماشین به منظور تخمین حداکثر عمق آبشستگی اطراف سازه باندا لایک می‌باشد. برای این منظور از روش‌های GEP، SVM و RF به همراه نتایج اطلاعات آزمایشگاهی استفاده

شده است. همچنین دقت مدل‌ها با ترکیب‌های متفاوت از ورودی‌ها مورد ارزیابی قرار گرفت و میزان اهمیت هر کدام از متغیرهای تاثیرگذار تعیین شد.

## مواد و روش‌ها

### جمع‌آوری اطلاعات

در تحقیق حاضر از نتایج یک تحقیق آزمایشگاهی انجام شده روی آبشکن نوع باندا لایک در آزمایشگاه هیدرولیک دانشکده مهندسی آب و محیط زیست دانشگاه شهید چمران اهواز استفاده شد (شجاعیان و همکاران، ۱۳۹۴). آزمایش‌ها در تحقیق مذکور با استفاده از یک فلوم آزمایشگاهی به طول ۷/۳۰ متر، عرض ۰/۵۶ متر و عمق ۰/۶۰ متر با شیب ثابت ۰/۰۰۲۸ انجام شده است. در انجام هر آزمایش از یک سری ۴ عددی آبشکن نوع باندا لایک استفاده شد. متغیرهای بدون بعد تاثیرگذار (متغیرهای ورودی) روی آبشستگی اطراف دماغه آبشکن عدد فرود ( $Fr$ )، زاویه نصب آبشکن نسبت به دیواره فلوم ( $\theta$ )، درصد نفوذپذیری قسمت متخلخل ( $\alpha$ ) و نسبت فاصله به طول آبشکن ( $S/L$ ) در نظر گرفته شد. لازم به ذکر است این متغیرها پس از انجام آنالیز ابعادی تعیین شدند. همچنین نسبت حداکثر عمق چاله آبشستگی نسبت به عمق جریان ( $h_s/h$ ) به عنوان متغیر خروجی در نظر گرفته شد (شکل ۲).



شکل ۲. متغیرهای ورودی به مدل‌های یادگیری ماشین موردنظر (شجاعیان و همکاران، ۱۳۹۴)

با توجه به تعداد در نظر گرفته شده در تحقیق مذکور برای هر متغیر در مجموع تعداد ۱۰۸ سری اطلاعات استخراج شد. ۸۰ درصد از سری اطلاعات (۸۶ سری) به منظور آموزش مدل‌ها و ۲۰ درصد باقیمانده (۲۲ سری) برای ارزیابی مدل‌ها استفاده شدند. جدول ۱ شاخص‌های آماری اطلاعات مورد استفاده در تحقیق حاضر را نشان می‌دهد.

جدول ۱. شاخص‌های آماری اطلاعات مورد استفاده

شاخص‌های آماری	متغیرهای ورودی				متغیر خروجی
	$Fr$	$\alpha$	$\theta$	$S/L$	$h_s/h$
Min	۰/۱۸	۰/۳۰	۶۰	۲	۰/۰۱
Max	۰/۳۴	۰/۶۴	۱۲۰	۳	۰/۵۴
Mean	۰/۲۱	۰/۴۶	۹۰	۲/۵	۰/۲۳
Std. deviation	۰/۰۲	۰/۱۴	۲۴/۰۶	۰/۴۱	۰/۱۳

ماشین بردار پشتیبان الگوریتمی است که نتایج موفقیت‌آمیزی را در زمینه‌های مختلف مهندسی آب، مانند تخمین شاخص کیفیت فاضلاب، پیش‌بینی جریان رودخانه، مدل‌سازی بارش- رواناب، تخمین سیل و خشکسالی و برآورد میزان رسوب نشان می‌دهد (Wang et al., 2018 ; Katipoğlu et al., 2023). SVM یک الگوریتم مبتنی بر کرنل می‌باشد که از روش‌های کاهش ریسک ساختاری و یادگیری آماری برای ایجاد ظرفیت بسط مناسب از طریق به حداقل رساندن خطای تعمیم در مقایسه با خطای آموزشی استفاده می‌کند. SVM با استفاده از یک تابع انتقال به نقشه‌برداری غیرخطی بردارهای ورودی در فضای ویژگی با ابعاد بالا کار می‌کند، که به کاهش پیچیدگی بهینه‌سازی کمک قابل توجهی می‌نماید (اسام و همکاران، ۲۰۲۲). روش ماشین بردار پشتیبان بر مبنای تعریف یک تقریب تابع رگرسیون بر اساس مجموعه‌ای از بردارهای پشتیبانی است که از یک مجموعه داده آموزشی نشأت می‌گیرد. با توجه به مطالعات موجود تابع SVM به شرح زیر بیان می‌شود (اسام و همکاران، ۲۰۲۲):

$$f(x) = \sum_{i=1}^N (\alpha_i - \alpha_i^*) K(x, z) + b_i \quad \text{رابطه (۱)}$$

که در آن  $(\alpha_i - \alpha_i^*)$  ضریب لاگرانژ،  $K(x, z)$  تابع کرنل و  $b_i$  مقدار خطای رابطه رگرسیونی می‌باشد. تابع کرنل نشان دهنده هاپی‌پارامتر اصلی SVM است که باید قبل از اجرای مدل انتخاب یا تنظیم شود. توابع کرنلی که می‌توان از آنها استفاده کرد تابع پایه شعاعی (RBF)، خطی، چند جمله‌ای و سیگموئید هستند. در مدل SVM انتخاب تابع کرنل بسیار مهم بوده و در مسایل مختلف با توجه به ماهیت مسئله می‌توان موارد متفاوتی را در نظر گرفت. انواع مختلف تابع کرنل در جدول ۲ ارائه شده است. تابع کرنل خطی، حالت خاصی از تابع کرنل چند جمله‌ای بوده و در مسایل ساده‌تر کارایی خوبی از خود نشان می‌دهد، ولی در مسایل پیچیده تابع کرنل چند جمله‌ای می‌تواند بسیار مفیدتر باشد. در تابع کرنل چند جمله‌ای پارامتر  $d$  باید از طرق آزمون و خطا یا از طریق الگوریتم‌های بهینه‌سازی محاسبه شود. تابع کرنل گوسی یا شعاعی (RBF) نوع دیگری از توابع کرنل بوده و معمولاً در مسائلی به کار برده می‌شود که هیچ گونه شناختی از نوع و ماهیت داده‌ها در دسترس نباشد. به طور کلی در توابع کرنل می‌باید مقادیر بهینه ضرایب  $c$ ،  $\epsilon$  و  $\gamma$  برآورد گردد. بدین منظور دو مشخصه  $\epsilon$  و  $C$  توسط الگوریتم بهینه‌سازی جستجوی شبکه<sup>۱</sup> و متغیر  $\gamma$  نیز به صورت آزمون و خطا محاسبه شد. پژوهش‌های پیشین تابع RBF را به عنوان بهترین تابع کرنل (به دلیل کارایی مناسب در بهینه‌سازی و سازگاری آن در حل مسائل پیچیده) پیشنهاد داده است (Tikhamarine et al., 2019; Parisouj et al., 2020; Essam et al., 2022). بنابراین در این پژوهش از تابع کرنل RBF برای انجام محاسبات استفاده شد.

جدول ۲. توابع کرنل رایج در ماشین‌های بردار پشتیبان (Hamel, 2009)

نوع تابع	تابع کرنل
خطی	$K(X_i, X_j) = X_i^T \times X_j$
چند جمله‌ای	$K(X_i, X_j) = (\gamma X_i^T \times X_j + C)^d$
تانژانت هیپربولیک	$K(X_i, X_j) = \tanh(\gamma X_i^T \times X_j + C)$

$$K(X_i, X_j) = -\exp(-\gamma |X_i - X_j|^2)$$

RFB

## مدل جنگل‌های تصادفی (RF)

الگوریتم جنگل تصادفی یا جنگل‌های تصادفی یکی از مدل‌های یادگیری ماشین می‌باشد که برای حل مسائل مختلف طبقه‌بندی و رگرسیون به کار گرفته می‌شود. این روش با تشکیل تعداد زیادی از درختان تصمیم‌گیر برای دریافت اطلاعات آموزشی اقدام و تعداد زیادی الگوریتم منفرد را با هم و با استفاده از قوانین حاکم، ترکیب می‌کند. اصل تکنیک آموزش گروهی بر این پایه استوار است که مشخصه دقت در این الگوریتم نسبت به سایر الگوریتم‌ها بیشتر است چون ترکیبی از چندین الگوریتم پیش‌بینی، دقت مناسب‌تری را برای مدل فراهم می‌کند (Breiman, 2001). در این روش درخت رگرسیونی (RT) مجموعه‌ای از شرایط و محدودیت‌ها را بیان می‌کند که به صورت سلسله مراتبی سازمان یافته‌اند و به صورت متوالی از گره ریشه به سمت بالا رشد کرده و به گره‌های برگ می‌رسند (Breiman, 2001). به منظور ایجاد درخت رگرسیونی از رگرسیون‌های چندگانه استفاده می‌گردد. عملیات تصمیم‌گیری در هر گره داخلی از گره ریشه، مطابق با قوانین درخت تکرار می‌گردد تا آن لحظه که شرایط توقف قبلی تعیین شده بدست آید. هر کدام از گره‌های آخری یا برگ‌ها به یک مدل رگرسیونی ساده که فقط در گره به کار برده می‌شود متصل می‌گردد. زمانی که فرایند فراخانی درخت به انتها نزدیک شود هرس کردن درختان یا پرونینگ برای افزایش راندمان ظرفیت تصمیم درخت‌ها توسط کاهش پیچیدگی ساختار بکار می‌رود. تعداد نمونه‌ها در گره‌ها می‌تواند به عنوان معیار پرونینگ در نظر گرفته شود (Breiman, 2001). مفهوم اصلی یک جنگل تصادفی این است که به طور مستقل چندین درخت تصمیم را بر روی زیر مجموعه‌های تصادفی مجموعه داده آموزشی اصلی بسازد. میانگین پیش‌بینی‌های این درخت‌های منفرد به منظور بهبود تعمیم‌پذیری/استحکام مدل مورد استفاده قرار می‌گیرند (He et al., 2022). به طور کلی می‌توان مراحل محاسباتی روش جنگل‌های تصادفی را به شرح زیر ارائه نمود (Kumar et al., 2023):

۱- انتخاب تصادفی زیر مجموعه: یک زیر مجموعه تصادفی از نقاط داده از مجموعه آموزشی انتخاب می‌شود. این زیر مجموعه معمولاً شامل کسری از کل نقاط داده است که با "p" نشان داده می‌شود.

۲- ساخت درخت تصمیم: با استفاده از زیر مجموعه نقاط انتخاب شده، درخت تصمیم ساخته می‌شود. این روش با استفاده از زیرمجموعه‌های مختلف داده برای کل درختان «N» تکرار می‌شود.

۳- تجمیع پیش‌بینی: هر یک از درخت‌های تصمیم «N» مقدار متغیر هدف را برای یک نقطه داده جدید پیش‌بینی می‌کند. نتایج تمام پیش‌بینی‌ها از درختان برای ارائه پیش‌بینی نهایی میانگین می‌شوند.

در روش RF بردار تصادفی  $X_n$  که مستقل از بردارهای تصادفی  $X_1, X_2, \dots, X_{n-1}$  بوده، برای درخت  $n$ ام، تولید می‌شود. همچنین همه بردارها از توزیع مشابهی تبعیت می‌کنند. رگرسیون درختی با استفاده از مجموعه داده‌های آموزش و  $X_n$  محاسبه شده مجموعه درختهایی برابر با  $n$  را به شرح زیر تولید می‌نماید (نوروزی و همکاران، ۱۳۹۶):

$$X_n = \{h_1(x), h_2(x), \dots, h_n(x)\}$$

رابطه ۲)

$$h_n = h(x, X_n), x = \{x_1, x_2, \dots, x_p\} \quad \text{رابطه ۳}$$

بردار  $p$  بعدی فوق یک جنگل را تشکیل داده و خروجی‌ها برای هر درخت به صورت زیر ارائه می‌شود:

$$y_1 = h_1(x), y_2 = h_2(x), \dots, y_n = h_n(x) \quad \text{رابطه ۴}$$

که در رابطه فوق  $y_n$  خروجی درخت  $n$ ام می‌باشد. برای به دست آوردن خروجی نهایی، متوسط همه پیش‌بینی‌های درخت‌ها محاسبه می‌شود (نوروزی و همکاران، ۱۳۹۶).

### روش برنامه‌ریزی بیان ژن (GEP)

روش برنامه‌ریزی بیان ژن (GEP) یک تکنیک جستجو و بهینه‌سازی اکتشافی است که از تکامل بیولوژیکی برای ایجاد راه حل‌های مناسب استفاده نموده و از آن برای برآورد یک رویداد خاص بهره می‌برد (Ferreira, 2022). در سال‌های اخیر، سیستم‌های متفاوتی از الگوریتم‌های ژنتیک گسترش یافته‌اند؛ الگوریتم‌های قدرتمندی که از سیر تکاملی طبیعی الهام گرفته‌اند و می‌توانند در طیف وسیعی از علوم بکار گرفته شوند. این الگوریتم‌ها از لحاظ ساختاری می‌توانند به سه گروه اصلی تقسیم شود (Ferreira, 2022): ۱) الگوریتم‌های ژنتیک با افرادی شامل کروموزوم‌های خطی با طول ثابت و بدون بیانی پیچیده. در این سیستم‌ها کروموزوم‌ها طبق برتری خصوصیات فردیشان حفظ می‌شوند و به الگوریتم ژنتیک<sup>۲</sup> (GA) معروف هستند. ۲) الگوریتم‌های ژنتیک با افرادی شامل ساختارهای شاخه‌ای<sup>۳</sup> از اندازه‌ها و اشکال متفاوت و توانا در پذیرش تعداد بیشتری از عوامل. در این سیستم‌ها ساختارهای شاخه‌ای طبق برتری خصوصیات فردیشان حفظ می‌شوند که برنامه‌ریزی ژنتیک (GP) نامیده می‌شوند. ۳) الگوریتم‌های ژنتیک با افراد کدگذاری شده به شکل کروموزوم‌های خطی با طولی ثابت و قابل بیان به شکل ساختارهای شاخه‌ای با اندازه‌ها و اشکال متفاوت. در این سیستم‌ها کروموزوم‌ها بواسطه برتری عوامل سببی روی فنوتیپ<sup>۴</sup> (ساختارهای شاخه‌ای) حفظ می‌شوند و به برنامه‌ریزی بیان ژن (GEP) معروف هستند (احمدی و همکاران، ۲۰۲۲).

یکی از مزایای GEP چند ژنی بودن آن است که امکان ایجاد برنامه‌های پیچیده‌تر با زیربرنامه‌های زیاد را فراهم می‌کند. در زیر مراحل کلی GEP ارائه می‌شود: مرحله ۱: یک جمعیت با انتخاب تصادفی کروموزوم‌ها (راه‌حل‌های مشکل بالقوه) ایجاد می‌شود. مرحله ۲: تعیین سازگاری هر کروموزوم ( $x$ ) با استفاده از تابع تناسب  $[f(x)]$ . مرحله ۳: مراحل زیر برای ایجاد مراحل جدید تکرار می‌شوند. (i) (انتخاب): بر اساس سازگاری آنها، دو کروموزوم والد از جمعیت انتخاب می‌شوند. (ii) (مقاطع): کروموزوم‌های والدین مرحله ۱ با احتمال خاصی تلاقی می‌کنند و دو فرزند ایجاد می‌کنند. (iii) (جهش): کروموزوم‌های کودک‌کان ایجاد شده با یک نرخ تصادفی با احتمال معین جهش می‌یابند. (IV) کودک‌کانی که در نتیجه عملگرهای ژنتیکی (انتخاب، مقاطع و جهش) ایجاد می‌شوند به جمعیت جدید اضافه می‌شوند. مرحله چهارم: در نسل جدید والدین جمعیت ایجاد شده جایگزین می‌شوند. مرحله ۵: الگوریتم خاتمه می‌یابد و در صورت رسیدن به شرایط مطلوب مانند دقت مورد نظر یا تعداد تکرارهای ذکر شده در مسئله، جمعیت موجود پاسخ مورد نظر را نشان می‌دهد (Azimi et al., 2017). مرحله ۶: اگر الگوریتم به مرحله پنج ختم نشود، مراحل دو تا شش برای رسیدن به نتایج مورد نظر تکرار می‌شود. کد ریاضی یک ژن را می‌توان به دو زبان بیان کرد: زبان ژن‌ها و زبان درختان بیان

2. Genetic Algorithm  
3. Ramified Structures  
4. Phenotype

(ET). کروموزومها و درختان بیان دو جزء اصلی GEP هستند. ژنهای GEP به دو بخش سر و دم تقسیم می‌شوند. سر شامل عملگرهای ریاضی است که برای نمایش یک عبارت ریاضی استفاده می‌شود و ثابت‌ها به عنوان نمادهای پایانی در دم گنجانده شده‌اند. اگر نمادهای پایانی در سر برای توضیح یک معادله ریاضی کافی نباشد، از نمادهای اضافی استفاده می‌شود. با شروع از خط بالای درخت و خواندن از چپ به راست، از بالا به پایین، ترجمه درخت بیان انجام می‌شود (Bagatur and Onen, 2016).

## الگوهای ورودی

برای دستیابی به بهترین عملکرد مدل‌ها نیاز است تا بهترین داده‌های آموزشی در فرآیند مدل‌سازی دخالت داده شود. این امر موجب می‌گردد ماهیت ساز و کار حاکم بر پدیده مورد بررسی و اطلاعات مورد نیاز مدل در اختیار آن قرار گرفته و نتایج برآوردها از اطمینان بیشتری برخوردار باشد. از طرفی افزایش تعداد ورودی‌ها نه تنها سبب پیچیدگی الگو و افزایش حافظه درگیر خواهد شد، بلکه سبب کاهش دقت نیز می‌گردد. بنابراین در تخمین حداکثر عمق ابستگي در دماغه  $(\frac{h_s}{h})$  بایستی سعی نمود موثرترین داده‌های اندازه‌گیری را به عنوان داده‌های آموزشی انتخاب کرد. بدین منظور در مطالعه حاضر متغیرهای مستقل  $(Fr, \frac{S}{L}, \alpha, \theta)$  در آزمایشگاه اندازه‌گیری و به مدل‌ها معرفی گردید. سناریوها بر این اساس تعریف شدند که حذف هر کدام از آنها به عنوان متغیر ورودی به چه میزان عملکرد مدل‌های یادگیری ماشین موردنظر بر اساس شاخص‌های آماری را تحت تاثیر قرار می‌دهد. در نتیجه سناریوهای ورودی یا به صورت تک متغیره هستند  $(S_1-S_4)$  و یا به صورت ترکیب‌های دو متغیره  $(S_5-S_{10})$  و سه متغیره  $(S_{11}-S_{14})$  می‌باشند. در نهایت یک سناریو هم به صورت ۴ متغیره  $(S_{15})$  تعریف شد. با استفاده از پارامترهای اندازه‌گیری شده در کل ۱۵ الگوی مختلف ورودی مطابق با جدول ۳ تعریف گردید.

جدول ۳. الگوهای مختلف ورودی به مدل‌ها

سناریو	داده‌های ورودی				خروجی مدل
	Fr	$\alpha$	S/L	$\theta$	
S <sub>1</sub>	✓				$h_s/h$
S <sub>2</sub>		✓			$h_s/h$
S <sub>3</sub>			✓		$h_s/h$
S <sub>4</sub>				✓	$h_s/h$
S <sub>5</sub>	✓	✓			$h_s/h$
S <sub>6</sub>	✓		✓		$h_s/h$
S <sub>7</sub>	✓			✓	$h_s/h$
S <sub>8</sub>		✓	✓		$h_s/h$
S <sub>9</sub>		✓		✓	$h_s/h$
S <sub>10</sub>			✓	✓	$h_s/h$
S <sub>11</sub>	✓	✓	✓		$h_s/h$
S <sub>12</sub>	✓	✓		✓	$h_s/h$
S <sub>13</sub>	✓		✓	✓	$h_s/h$
S <sub>14</sub>		✓	✓	✓	$h_s/h$
S <sub>15</sub>	✓	✓	✓	✓	$h_s/h$



## ارزیابی مدل‌ها

در این تحقیق، برای ارزیابی مدل‌های موردنظر از معیارهای جذر میانگین مربعات خطا (RMSE)، میانگین خطای مطلق (MAE) و ضریب همبستگی استفاده شد. شاخص ارزیابی RMSE نشان دهنده‌ی میزان خطای پیش‌بینی در فرایند مدل‌سازی است. هر چقدر مقادیر RMSE کم‌تر و به سمت عدد صفر حرکت کند نشانگر این است که مدل از دقت بالایی برخوردار می‌باشد. میانگین خطای مطلق، مقدار میانگین مثبت خطاها در یک مجموعه از مقادیر پیش‌بینی شده را محاسبه می‌کند و رفتار آن مشابه با شاخص RMSE می‌باشد. در ادامه روابط شاخص‌های مورد استفاده ارائه شده است :

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (P_i - O_i)^2}{n}} \quad \text{رابطه ۵}$$

$$MAE = \frac{\sum_{i=1}^n |P_i - O_i|}{n} \quad \text{رابطه ۶}$$

$$CC = \left( 1 - \frac{\sum_{i=1}^n (P_i - O_i)^2}{\sum_{i=1}^n (P_i - \bar{O})^2} \right)^{0.5} \quad \text{رابطه ۷}$$

که در روابط فوق  $O_i$  مقدار اندازه‌گیری شده در گام زمانی  $i$  ام،  $P_i$  مقدار محاسبه شده در همان زمان،  $n$  تعداد داده‌ها و  $\bar{O}$  میانگین مقادیر اندازه‌گیری شده می‌باشد. لازم به ذکر است در مقالات متعددی از این شاخص‌ها به منظور ارزیابی مدل‌ها استفاده شده است (Hassanzadeh and Abbaszadeh, 2023; Abbaszadeh et al., 2023; Abbaszadeh et al., 2024; دانش‌فراز و همکاران، ۱۴۰۰).

## نتایج و بحث

همان‌طور که اشاره شد در این پژوهش مدل‌سازی حداکثر عمق آبخستگی اطراف آبشکن باندا لایک با استفاده از روش‌های برنامه ریزی بیان ژن (GEP)، ماشین بردار پشتیبان (SVM) و جنگل‌های تصادفی (RF) انجام شد. بدین منظور پارامترهای درصد نفوذپذیری آبشکن ( $\theta$ )، فاصله آبشکن‌ها نسبت به همدیگر (S/L)، درصد بازشدگی ( $\alpha$ ) و عدد فرود (Fr) به عنوان متغیرهای اصلی در نظر گرفته شد. همچنین در پژوهش حاضر ۸۰ درصد داده‌ها به عنوان داده‌های آموزشی و ۲۰ درصد برای صحت‌سنجی عملکرد مدل‌ها در پیش‌بینی عمق آبخستگی اطراف آبشکن باندا لایک استفاده شد.

نتایج حاصل از اجرای مدل GEP در جدول ۴ ارائه شده است. با توجه به این جدول مشاهده می‌شود که در سناریوهایی که فقط یک ورودی برای مدل‌سازی عمق آبخستگی مورد استفاده قرار گرفته است (سناریوهای S1 تا S4) کمترین خطا مربوط به

الگوی S2 با پارامتر  $\alpha$  می باشد. این بدان مفهوم است که در مقایسه با سایر پارامترهای اندازه گیری شده متغیر  $\alpha$  بیشترین S/L و کمترین تاثیر را در بهبود عملکرد مدل GEP براساس شاخص CC داشته است.

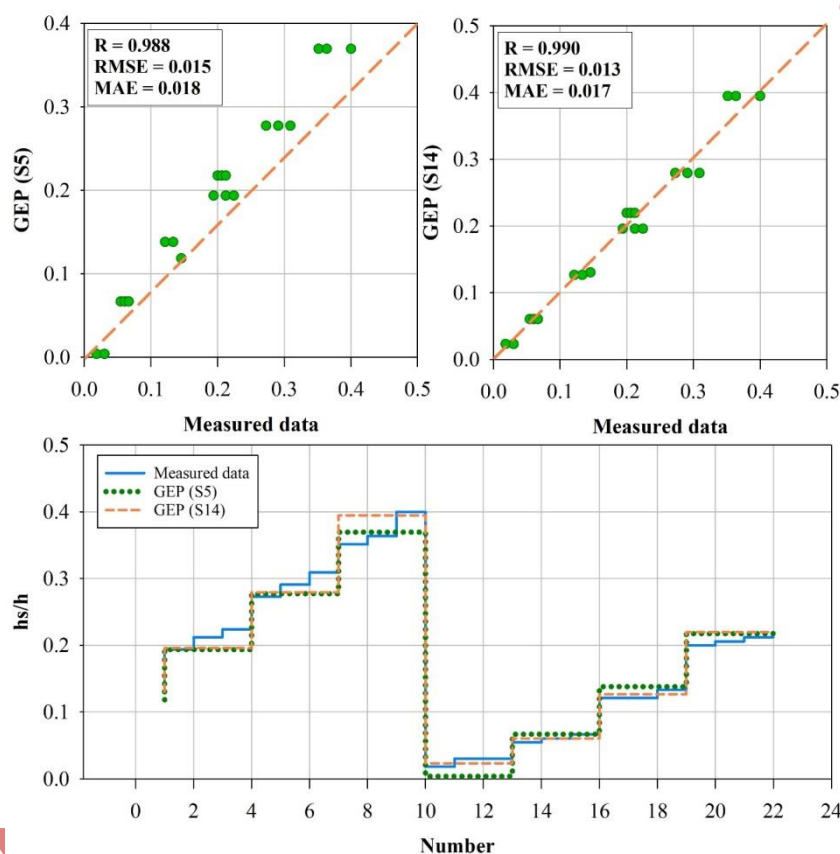
جدول ۴. نتایج شاخص های آماری ارزیابی عملکرد برای روش GEP در مدل سازی عمق آبستگي اطراف آبشكن باندال لایک

سناریو	Train			test		
	RMSE	MAE	CC	RMSE	MAE	CC
S1	۰/۱۰	۰/۰۹۰	۰/۶۹	۰/۰۸۵	۰/۱۰۷	۰/۷۲۸
S2	۰/۱۰	۰/۰۸۰	۰/۶۹	۰/۰۶۶	۰/۰۷۷	۰/۷۶۱
S3	۰/۱۴	۰/۱۱۰	۰/۲۴	۰/۱۴۷	۰/۱۷۳	۰/۷۰۱
S4	۰/۱۴	۰/۱۲۰	۰/۰۶	۰/۱۰۹	۰/۱۲۸	۰/۰۶۲
S5	۰/۰۳	۰/۰۲۰	۰/۹۸	۰/۰۱۵	۰/۰۱۸	۰/۹۸۸
S6	۰/۰۹	۰/۰۸۰	۰/۷۴	۰/۱۵۳	۰/۱۷۷	۰/۷۲۵
S7	۰/۱۰	۰/۰۹۰	۰/۶۹	۰/۰۸۵	۰/۱۰۷	۰/۷۳۱
S8	۰/۱۰	۰/۰۸۰	۰/۶۹	۰/۰۶۶	۰/۰۷۶	۰/۷۵۶
S9	۰/۱۴	۰/۱۱۰	۰/۲۵	۰/۱۴۹	۰/۱۷۵	۰/۰۶۲
S10	۰/۱۰	۰/۰۸۰	۰/۷۰	۰/۰۷۳	۰/۰۸۸	۰/۷۶۱
S11	۰/۰۳	۰/۰۲۰	۰/۹۸	۰/۰۱۹	۰/۰۲۲	۰/۹۹۲
S12	۰/۱۲	۰/۰۹۰	۰/۶۰	۰/۰۷۷	۰/۰۸۷	۰/۷۳۵
S13	۰/۱۰	۰/۰۸۰	۰/۷۱	۰/۰۸۷	۰/۱۰۶	۰/۷۲۳
S14	۰/۰۲	۰/۰۲۰	۰/۹۹	۰/۰۱۳	۰/۰۱۷	۰/۹۹۰
S15	۰/۰۲	۰/۰۲۱	۰/۹۷	۰/۰۲۱	۰/۰۲۵	۰/۹۷۶

همچنین با توجه به جدول ۴ مشاهده می شود که با افزایش تعداد ورودی ها عملکرد مدل GEP به طور معنی داری بهبود می یابد به طوریکه در الگوهای ورودی با دو متغیر (S5-S10) دقت پیش بینی ها نسبت به الگوهای تک متغیره (S1-S4) بهبودی ۱۱ درصدی را در شاخص RMSE نشان می دهد. همین روند در الگوهای سه متغیره نیز تکرار می گردد.

در سناریوهایی که تأثیر پارامترها به صورت دو ورودی بررسی شده اند، مدل GEP با الگوی S5 کمترین میزان شاخص MAE (برابر با ۰/۰۱۸) را در پیش بینی میزان آبستگي به خود اختصاص داده است. این الگو متغیرهای  $\alpha$  و عدد Fr را شامل می شود. در خصوص سناریوها با ترکیب سه متغیر (S11-S14) بیشترین دقت مربوط به الگوی S14 با پارامترهای  $\alpha$ ، Fr،  $\theta$  بوده و شاخص های

آماري RMSE، MAE و CC به ترتيب برابر با ۰/۰۱۷، ۰/۰۱۳ و ۰/۹۹۰ حاصل شده است. در الگوي S15 تمام پارامترهاي مورد بررسي در پيش بيني ميزان آبشستگي دخالت داده شدند. نتايج نشان داد که با دخالت کليه پارامترها دقت عملکرد مدل GEP نسبت به زماني که از سه متغير  $\alpha$ ، Fr، S/L يا همان الگوي S14 کاهش يافته و شاخص هاي خطا در پيش بيني عمق آبشستگي اندکي افزايش يافته است. به عبارت ديگر از بين کل الگوها، الگوي S14 بيشترين دقت را براي مدل GEP فراهم مي کند. چنانچه آستانه خطا اندکي افزايش يابد، مي توان تعداد پارامترهاي وردي را کاهش داده و الگوي S5 را نيز در پيش بيني عمق آبشستگي آبشکن باندا لايک با استفاده از مدل GEP توصيه نمود. در شکل ۳ نمودارهاي پراکندي و گرافيکي مربوط به داده هاي محاسباتي و مشاهداتي براي سناريوهاي S5 و S14 ارائه شده است.



شکل ۳. نمودارهاي گرافيکي و پراکندي حاصل از عملکرد مدل GEP در برآورد حداکثر عمق آبشستگي اطراف آبشکن نوع باندا لايک در مرحله آزمون

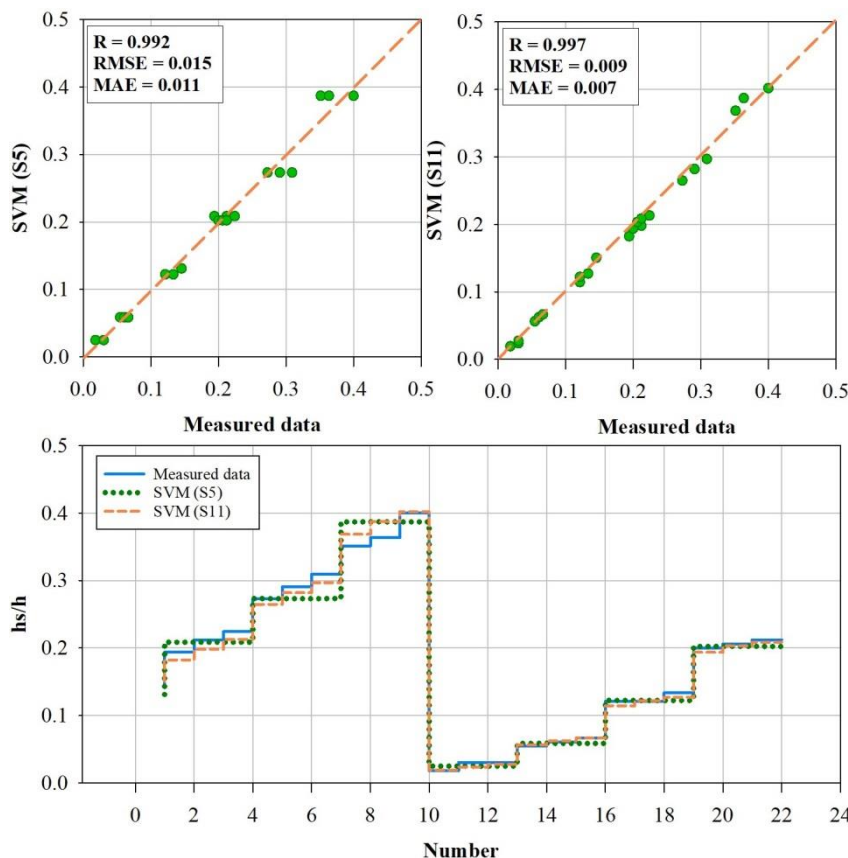
در جدول ۵ شاخص هاي آماري ارزيابي عملکرد مدل SVM در هر دو مرحله آموزش و آزمون ارائه شده است. همانگونه که مشاهده مي شود پارامتر  $\alpha$  بيشترين دقت را براي اين مدل فراهم نموده است. به عبارت ديگر با مقايسه نتايج الگوهاي تک وردي بين SVM و GEP مي توان عملکرد يکسان را مشاهده نمود اما خطا در روش SVM افزايش داشته است به طوريکه در سناريوهاي S1 تا S4 ميانگين شاخص RMSE ۱۴ درصد رشد را نشان مي دهد.

جدول ۵. نتایج شاخص های آماری ارزیابی عملکرد برای روش SVM در مدل سازی عمق آبستگي اطراف آبشکن باندال لایک

Pattern	Train			test		
	RMSE	MAE	CC	RMSE	MAE	CC
S1	۰/۰۹	۰/۱۰	۰/۶۸	۰/۱۰۷	۰/۰۸۵	۰/۷۳۲
S2	۰/۰۸	۰/۱۰	۰/۶۹	۰/۰۷۸	۰/۰۶۷	۰/۷۶۱
S3	۰/۱۱	۰/۱۴	۰/۲۲	۰/۱۵۰	۰/۱۲۷	۰/۰۰۰
S4	۰/۱۱	۰/۱۲	۰/۲۰	۰/۱۲۸	۰/۱۰۹	۰/۰۰۰
S5	۰/۰۲	۰/۰۲	۰/۹۹	۰/۰۱۵	۰/۰۱۱	۰/۹۹۲
S6	۰/۰۸	۰/۰۹	۰/۷۴	۰/۱۹۰	۰/۱۷۱	۰/۷۳۲
S7	۰/۰۹	۰/۱۰	۰/۶۸	۰/۱۰۶	۰/۰۸۱	۰/۷۴۰
S8	۰/۰۸	۰/۱۰	۰/۷۰	۰/۰۷۸	۰/۰۶۷	۰/۷۵۵
S9	۰/۱۱	۰/۱۴	۰/۲۲	۰/۱۴۹	۰/۱۲۷	۰/۰۶۲
S10	۰/۰۸	۰/۱۰	۰/۷۱	۰/۲۱۰	۰/۱۹۲	۰/۷۶۱
S11	۰/۰۲	۰/۰۲	۰/۹۹	۰/۰۰۹	۰/۰۰۷	۰/۹۹۷
S12	۰/۰۸	۰/۰۹	۰/۷۴	۰/۱۸۴	۰/۱۶۷	۰/۷۴۳
S13	۰/۰۸	۰/۱۰	۰/۷۱	۰/۱۵۹	۰/۱۴۲	۰/۷۵۹
S14	۰/۰۱	۰/۰۱	۰/۹۹	۰/۱۳۲	۰/۱۱۴	-۰/۱۹۵
S15	۰/۰۱	۰/۰۱	۰/۹۷	۰/۳۲۹	۰/۲۶۰	-۰/۰۹۸

افزایش تعداد ورودی ها در مدل SVM همچون مدل GEP لزوماً به معنای بهبود عملکرد نبوده و رفتاری متفاوت را نشان می دهد. به طوریکه با الگوهای دو متغیره میانگین شاخص MAE نسبت به الگوهای تک متغیره افزایشی حدوداً دو برابری را نشان می دهد اما الگوهای سه متغیره کاهش ۱۵ درصدی را در این شاخص گزارش می دهد. در سناریوهایی که تأثیر پارامترها به صورت ورودی های چندگانه بررسی شده اند، مدل SVM با الگوی های S5 و S11 کمترین میزان شاخص MAE (به ترتیب برابر با ۰/۰۱۱ و ۰/۰۰۷) را در پیش بینی میزان آبستگي به خود اختصاص داده است. این الگوها متغیرهای  $\alpha$ ، عدد Fr و  $\theta$  را شامل می شود. در الگوهایی که از سه و چهار پارامتر استفاده می شود (S11 تا S15) عملکرد مدل SVM به شدت کاهش داشته و افزایش تعداد ورودی ها نتوانسته برای این روش مفید باشد. به عبارت دیگر مدل SVM برای دستیابی به بهترین دقت به تعداد پارامترهای بیشتری نیاز ندارد بلکه به ترکیب صحیح ورودی ها حساسیت بیشتری نشان می دهد. با مقایسه میانگین شاخص های خطا مشاهده می شود که مدل SVM خطای بیشتری را نسبت به روش GEP داشته اما با ترکیب ورودی های مناسب در الگوی S11 نتوانسته

خطای کمتر و دقت بسیار بهتری را با مقادیر RMSE برابر با ۰/۰۰۹، MAE برابر با ۰/۰۰۷ و ضریب همبستگی معادل با ۰/۹۹۷ ارائه دهد. در شکل ۴ نمودارهای پراکندگی و گرافیکی مربوط به داده های محاسباتی و مشاهداتی برای سناریوهای S5 و S11 برای مدل SVM ارائه شده است.



شکل ۴. نمودارهای گرافیکی و پراکندگی حاصل از عملکرد مدل SVM در برآورد حداکثر عمق آبخستگی در اطراف آبشکن نوع باندا لایک در مرحله آزمون

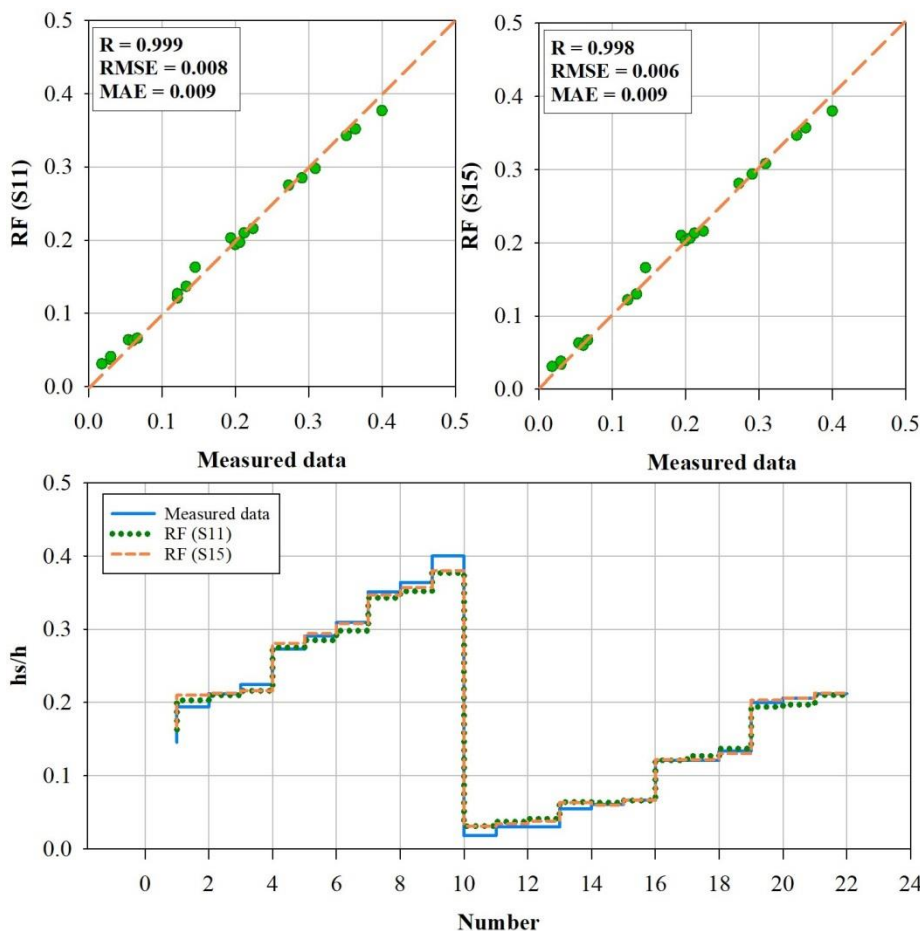
جدول ۶ شاخص های آماری ارزیابی عملکرد مدل RF را برای هر دو مرحله آموزش و آزمون نشان می دهد. همانطور که مشاهده می شود در سناریوهای انفرادی که شامل یک پارامتر ورودی می باشد (S1-S4)، مؤثرترین پارامتر در پیش بینی حداکثر عمق آبخستگی اطراف دماغه ی باندا لایک، پارامتر درصد بازشدگی و مربوط به سناریو S2 می باشد. شاخص های آماری مربوط به این سناریو شامل RMSE، MAE و CC به ترتیب برابر با ۰/۰۶۴، ۰/۰۶۷ و ۰/۷۱۶۱ می باشد. همچنین پارامتر  $\theta$  مربوط به سناریو S3 به عنوان کم اهمیت ترین متغیر شناخته شد.

جدول ۶. نتایج شاخص های آماری ارزیابی عملکرد برای روش RF در مدل سازی عمق آبخستگی اطراف آبشکن باندا لایک

Pattern	Train			test		
	RMSE	MAE	CC	RMSE	MAE	CC
S1	۰/۰۹	۰/۰۱	۰/۷۴۸	۰/۰۷۴	۰/۰۷۷	۰/۷۳۱
S2	۰/۰۸	۰/۰۱	۰/۷۸۹	۰/۰۶۴	۰/۰۷۳	۰/۷۶۱
S3	۰/۱۱	۰/۱۴	۰/۰۰۰	۰/۰۹۵	۰/۱۱۳	۰/۰۰۰
S4	۰/۱۱	۰/۱۴	۰/۰۷۸	۰/۰۹۶	۰/۱۱۲	۰/۰۶۲

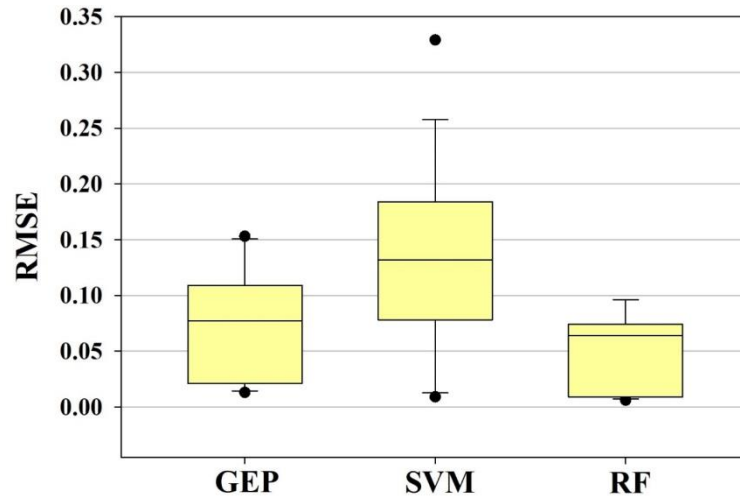
S5	۰/۰۲	۰/۰۲	۰/۹۹۹	۰/۰۰۹	۰/۰۱۳	۰/۹۹۴
S6	۰/۰۸	۰/۰۹	۰/۸۳۱	۰/۰۷۴	۰/۰۷۷	۰/۷۳۱
S7	۰/۰۹	۰/۰۱	۰/۸۳۰	۰/۰۷۳	۰/۰۷۶	۰/۷۴۰
S8	۰/۰۸	۰/۰۱	۰/۸۲۱	۰/۰۶۳	۰/۰۷۳	۰/۷۶۲
S9	۰/۱۱	۰/۱۴	۰/۰۷۹	۰/۰۹۶	۰/۱۱۲	۰/۰۶۲
S10	۰/۰۸	۰/۰۱	۰/۸۱۲	۰/۰۶۴	۰/۰۷۳	۰/۷۶۱
S11	۰/۰۲	۰/۰۲	۰/۹۹۹	۰/۰۰۸	۰/۰۰۹	۰/۹۹۹
S12	۰/۰۸	۰/۰۹	۰/۸۲۱	۰/۰۷۳	۰/۰۷۶	۰/۷۳۹
S13	۰/۰۸	۰/۱۰	۰/۸۹۲	۰/۰۶۳	۰/۰۷۳	۰/۷۶۲
S14	۰/۰۱	۰/۰۱	۰/۹۹۶	۰/۰۰۹	۰/۰۱۲	۰/۹۹۵
S15	۰/۰۰	۰/۰۱	۰/۹۹۹	۰/۰۰۶	۰/۰۰۹	۰/۹۹۸

در خصوص الگوهایی که پارامترهای ورودی به صورت دوتایی مورد بررسی قرار گرفت، سناریوی S5 با دارا بودن پارامترهای عدد فرود و درصد بازشدگی ( $\alpha$ , Fr) بیشترین دقت را در پیش‌بینی حداکثر عمق آبستگي در میان سناریوهای دو تایی دارا بود. در این سناریو میزان شاخص‌های آماری میانگین مجذور مربعات خطا، میانگین قدر مطلق خطا و ضریب همبستگی به ترتیب برابر با ۰/۰۰۹، ۰/۰۱۳ و ۰/۹۹۴ به دست آمد. در خصوص ترکیب‌های سه تایی از پارامترهای موجود در سناریو S11 شامل پارامترهای ( $\alpha$ , S/L, Fr) نیز MAE, RMSE و CC به ترتیب معادل ۰/۰۰۸، ۰/۰۰۹ و ۰/۹۹۹ تعیین شدند. سناریو ۱۵ که شامل تمام پارامترها به عنوان داده‌های ورودی به مدل RF بوده خطایی کمتر از الگوهای قبلی ارائه کرده است. این امر نشان می‌دهد که مدل RF برای دست‌یابی به بهترین عملکرد خود نیاز به پارامترهای بیشتری نسبت به روش‌های SVM و GEP دارد. اما نکته مهمی که باید مورد توجه قرار گیرد آن است که در روش‌های SVM و GEP افزایش تعداد ورودی‌ها لزوماً باعث بهبود نتایج نمی‌شد و بعضاً الگوهای دو یا سه پارامتری خطای بزرگی را نشان می‌دادند. این در حالی است که نوسان خطا در روش RF با الگوهای دو، سه و چهار پارامتری بسیار کمتر می‌باشد. از بین کل سناریوها سناریوهای S11 و S15 به ترتیب بیشترین دقت و کمترین خطا را به خود اختصاص داده‌اند. الگوی S5 نیز می‌تواند با افزایش اندکی در خطای قابل قبول به عنوان الگویی که اطلاعات کمتری نیاز دارد به عنوان الگوی ارزیابی اولیه انتخاب گردد. در شکل ۵ نمودارهای پراکندگی و گرافیکی مربوط به داده‌های محاسباتی و مشاهداتی برای سناریوهای S11 و S15 برای مدل SVM ارائه شده است.



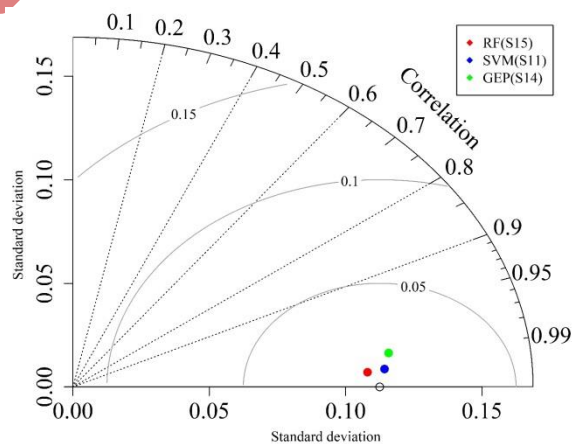
شکل ۵. نمودارهای گرافیکی و پراکندگی حاصل از عملکرد مدل RF در برآورد حداکثر عمق آبستگي در اطراف آبشکن نوع باندا لایک در مرحله آزمون

در شکل ۶ نمودار جعبه و خط (باکس و ویسکر) RMSE، برای مدل‌های GEP، SVM و RF در مرحله آزمون ارائه شده است. در این شکل خط رسم شده در داخل مستطیل‌ها، میانه RMSE الگوهای مختلف ورودی می‌باشد. عرض مستطیل‌ها در قسمت فوقانی، صدک ۷۵ و در قسمت تحتانی صدک ۲۵ را نشان می‌دهد. همچنین قسمت‌های انتهایی خطوط قائم در قسمت فوقانی و تحتانی به ترتیب نشان دهنده حداکثر و حداقل مقدار خطا برای الگوهای مختلف ورودی مدل‌های مورد بررسی می‌باشد. با توجه به شکل ۶ مشاهده می‌شود که عملکرد مدل RF در مرحله صحت سنجی بسیار بهتر از روش‌های SVM و GEP می‌باشد. همچنین پراکندگی خطا در روش SVM بسیار بیشتر از سایر روش‌ها می‌باشد. با مقایسه نمودار جعبه ای سه روش می‌توان نتیجه گرفت که روش RF در برآورد عمق آبستگي آبشکن نوع باندا لایک از عملکرد بسیار بهتری برخوردار بوده و مدل با ورودی‌های یکسان از پراکنش خطای کمتری برخوردار است.



شکل ۶. نمودار جعبه ای شاخص RMSE برای مدل‌های GEP، SVM و RF در مرحله صحت سنجی

علاوه بر مقایسه عملکرد مدل‌های به کار گرفته شده با استفاده از کلیه سناریوها، بهترین حالت این مدل‌ها با الگوهای ورودی بهینه نیز از طریق نمودار تیلور مورد بررسی و ارزیابی قرار گرفت (شکل ۷). نمودار تیلور به دو صورت نیم‌دایره برای نمایش همبستگی منفی و مثبت به صورت همزمان یا ربع‌دایره فقط برای نمایش همبستگی مثبت قابل ترسیم می‌باشد. در هر دو صورت، مقادیر ضریب همبستگی روی قوس آن و به صورت شعاع دایره بوده و مقادیر انحراف معیار در محورهای عمودی و افقی واقع می‌شود. علاوه بر این در نمودار تیلور یک نقطه مرجع تعریف می‌گردد که همان انحراف از معیار سری داده‌های مشاهده‌ای است. هرچقدر موقعیت سری داده‌های شبیه‌سازی شده توسط هر یک از مدل‌ها به نقطه مرجع نزدیک‌تر باشد، عملکرد آن مدل بهتر ارزیابی می‌شود (Dehghani et al., 2020). بنابراین با توجه به نمودار تیلور رسم شده برای مدل‌های مبتنی بر یادگیری ماشین در برآورد حداکثر عمق آبستگي اطراف دماغه آبسکن نوع باندا لایک، مشاهده می‌شود که روش RF با سناریوی S15 کمترین فاصله از نقطه مرجع و بیشترین ضریب همبستگی را به خود اختصاص داده و به عنوان گزینه برتر انتخاب می‌شود. همچنین این شکل نشان می‌دهد که عملکرد سایر روش‌ها نیز به یکدیگر نزدیک می‌باشد.



شکل ۷. نمودار تیلور برای مدل‌های GEP، SVM و RF در مرحله صحت سنجی



## نتیجه گیری

همانطور که عنوان شد پیش بینی حداکثر عمق آبستگی دماغه آبشکن به منظور در نظر گرفتن تمهیداتی در راستای جلوگیری از تخریب آن بسیار حائز اهمیت می باشد. لذا تا کنون روابط تجربی متعددی توسط محققین مختلف در خصوص آبشکن صلب ارائه شده است. در تحقیق حاضر دقت روش های یادگیری ماشین (RF, SVM, GEP) در خصوص پیش بینی حداکثر عمق آبستگی با در نظر گرفتن ترکیب های متفاوت از متغیرهای ورودی در دماغه آبشکن نوع باندال لایک مورد ارزیابی قرار گرفت. برای تمامی مدل ها برای سناریوهای تک متغیره بهترین عملکرد مربوط به سناریوی S2 با پارامتر ورودی  $\alpha$  تعیین شد. در این راستا شاخص MAE برای مدل های RF و SVM, GEP به ترتیب معادل ۰/۰۷۷، ۰/۰۶۷ و ۰/۰۷۳ برآورد گردید. همچنین پایین ترین سطح عملکرد مربوط به سناریوی S3 با پارامتر ورودی S/L تعیین شد. این بدان مفهوم است که در مقایسه با سایر پارامترهای اندازه گیری شده متغیر  $\alpha$  بیشترین و S/L کمترین تاثیر را در بهبود عملکرد مدل های مورد بررسی دارند. برای سناریوهای دو متغیره نتایج حاصل نشان داد که بهترین ترکیب ورودی برای تمامی مدل ها سناریوی S5 با متغیرهای  $\alpha$  و Fr می باشد. بر این اساس شاخص RMSE برای مدل های RF و SVM, GEP به ترتیب معادل ۰/۰۱۵، ۰/۰۱۵ و ۰/۰۰۹ تعیین گردید. با افزایش تعداد ورودی ها به دو متغیره، متوسط شاخص MAE برای مدل های RF و GEP به ترتیب ۱۲ درصد و ۲۴ درصد کاهش داشته که نشان دهنده افزایش دقت با افزایش تعداد ورودی برای این دو مدل می باشد. اما متوسط این شاخص برای مدل SVM افزایش ۱۲ درصدی داشته که بیانگر کاهش دقت ارزیابی این مدل با افزایش تعداد ورودی می باشد. در خصوص سناریوهای سه متغیره برای مدل GEP بیشترین دقت مربوط به سناریوی S14 با پارامترهای  $\alpha$ ، Fr،  $\theta$  بوده و شاخص های آماری RMSE، MAE و CC به ترتیب برابر با ۰/۰۱۳، ۰/۰۱۷ و ۰/۹۹۰ حاصل شد. همچنین میانگین شاخص MAE برای سناریوهای سه متغیره در مدل GEP نسبت به سناریوهای دو متغیره ۴۶ درصد کاهش داشت که همچنان بیانگر افزایش دقت این مدل با افزایش تعداد ورودی هاست. در مدل های RF و SVM بهترین سناریوی سه متغیره S11 با پارامترهای  $\alpha$ ، Fr، S/L بوده که شاخص ارزیابی RMSE برای این دو مدل به ترتیب معادل ۰/۰۰۹ و ۰/۰۰۸ تعیین شدند. مشخص شد مدل RF نیز مانند مدل GEP با افزایش تعداد متغیرهای ورودی افزایش دقت (کاهش ۴۱ درصدی میانگین شاخص MAE) داشته است. حال آنکه افزایش تعداد ورودی برای مدل SVM تغییر محسوسی در دقت آن ایجاد نکرده است. برای سناریوهای چهار متغیره (S15) مدل SVM با کاهش دقت شدیدی مواجه شد و میانگین شاخص MAE افزایش تقریبی ۱۴۰ درصد داشت. حال آنکه مدل های RF و GEP برای سناریوی S15 نیز افزایش دقت را تجربه کردند. هر چند میزان افزایش دقت مدل RF برای این سناریو بیشتر از مدل GEP حاصل شد. چرا که میانگین شاخص MAE برای مدل RF برای سناریوی S15، ۷۸ درصد کاهش و برای مدل GEP ۵۷ درصد کاهش را نشان داد. این امر نشان می دهد که مدل RF برای دست یابی به بهترین عملکرد خود نیاز به پارامترهای بیشتری نسبت به روش های SVM و GEP دارد.

## سپاسگزاری

در پایان از حمایت مالی شورای پژوهشی دانشگاه شهید چمران اهواز کمال تشکر را داریم (SCU.WH1402.31370).

## منابع

روشنگر، کیومرث؛ گودرزی، سپهر؛ عباس زاده، حمیدرضا (۱۴۰۲). بررسی عددی عملکرد آبشکنهای تیغه ای بر آبستگی و تأثیر آن بر پارامترهای هیدرولیکی رسوب و جریان. محیط زیست و مهندسی آب، ۱۰(۱)، ۱۳۶-۱۲۱.

شجاعیان، زهرا؛ کاشفی پور، سید محمود؛ موسوی جهرمی، سید حبیب؛ شفاعی بجستان، محمود (۱۳۹۴). مطالعه آزمایشگاهی آبستگي موضعی اطراف سری آبشکن‌های نوع باندا لایک در شرایط آب زلال. *علوم و مهندسی آبیاری*، ۳۸(۲)، ۳۲-۲۱.

دانشفراز، رسول؛ نوروزی، رضا؛ عباس‌زاده، حمیدرضا (۱۴۰۰). بررسی عددی پارامترهای مؤثر بر هیدرولیک جریان در سرریزهای تناسبی دودکشی، *تحقیقات آب و خاک ایران*، ۵۲(۶)، ۱۶۱۶-۱۵۹۹.

نوروزی، حسین؛ ندیری، عطاء الله؛ اصغری مقدم، اصغر و قره‌خانی، مریم (۱۳۹۶). پیش‌بینی قابلیت انتقال آبخوان دشت ملکان با استفاده از روش جنگل تصادفی. *دانش آب و خاک*، ۲۷(۲)، ۶۱-۷۵.

## References

- Abbaszadeh, H., Norouzi, R., Sume, V., Kuriqi, A., Daneshfaraz, R., & Abraham, J. (2023). Sill role effect on the flow characteristics (experimental and regression model analytical). *Fluids*, 8(8), 235.
- Abbaszadeh, H., Daneshfaraz, R., Sume, V., & Abraham, J. (2024). Experimental investigation and application of soft computing models for predicting flow energy loss in arc-shaped constrictions. *AQUA—Water Infrastructure, Ecosystems and Society*, jws2024010.
- Ahmadi, F., Mehdizadeh, S., Mohammadi, B., Pham, Q. B., Doan, T. N. C., & Vo, N. D. (2021). Application of an artificial intelligence technique enhanced with intelligent water drops for monthly reference evapotranspiration estimation. *Agricultural Water Management*, 244, 106622.
- Azimi, H., Shabanlou, S., Ebtehaj, I., Bonakdari, H., & Kardar, S. (2017). Combination of computational fluid dynamics, adaptive neuro-fuzzy inference system, and genetic algorithm for predicting discharge coefficient of rectangular side orifices. *Journal of Irrigation and Drainage Engineering*, 143(7), 04017015.
- Bagatur, T., & Onen, F. (2016). Computation of design coefficients in ogee-crested spillway structure using GEP and regression models. *KSCE Journal of Civil Engineering*, 20, 951-959.
- Basser, H., Karami, H., Shamshirband, S., Jahangirzadeh, A., Akib, S., & Saboohi, H. (2014). Predicting optimum parameters of a protective spur dike using soft computing methodologies—A comparative study. *Computers & Fluids*, 97, 168-176.
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), pp. 5-32.
- Daneshfaraz, R., Norouzi, R., & Abbaszadeh, H. (2021). Numerical Investigation on Effective Parameters on Hydraulic Flows in Chimney Proportional Weirs. *Iranian Journal of Soil and Water Research*, 52(6), 1599-1616 (In Persian).
- Dehghani, R., Torabi Poudeh, H., Younesi, H., & Shahinejad, B. (2020). Daily streamflow prediction using support vector machine-artificial flora (SVM-AF) hybrid model. *Acta Geophysica*, 68, 1763-1778.
- Essam, Y., Huang, Y. F., Ng, J. L., Birima, A. H., Ahmed, A. N., & El-Shafie, A. (2022). Predicting streamflow in Peninsular Malaysia using support vector machine and deep learning algorithms. *Scientific Reports*, 12(1), 3883.
- Ferreira, C. (2002). Genetic representation and genetic neutrality in gene expression programming. *Advances in Complex System*, (5)4, 389-408.
- Ferreira, C. Gene expression programming: A new adaptive algorithm for solving problems. arXiv 2001, arXiv:cs/0102027.
- Hassanzadeh, Y., & Abbaszadeh, H. (2023). Investigating Discharge Coefficient of Slide Gate-Sill Combination Using Expert Soft Computing Models. *Journal of Hydraulic Structures*, 9(1), 63-80.
- He, S., Wu, J., Wang, D., & He, X. (2022). Predictive modeling of groundwater nitrate pollution and evaluating its main impact factors using random forest. *Chemosphere*, 290, 133388.
- Katipoğlu, O. M., Yeşilyurt, S. N., Dalkılıç, H. Y., & Akar, F. (2023). Application of empirical mode decomposition, particle swarm optimization, and support vector machine methods to predict stream flows. *Environmental Monitoring and Assessment*, 195(9), 1108.
- Kumar, S., Goyal, M.K., Deshpande, V., & Agarwal, M. (2023). Estimation of time dependent scour depth around circular bridge piers: Application of ensemble machine learning methods. *Ocean Engineering*, 270, 113611.
- Kumar, V., Kedam, N., Sharma, K. V., Mehta, D. J., & Caloiero, T. (2023). Advanced machine learning

- techniques to improve hydrological prediction: a comparative analysis of streamflow prediction models. *Water*, 15(14), 2572.
- Najafzadeh, M., Oliveto, s. (2021). More reliable predictions of clear-water scour depth at pile groups by robust artificial intelligence techniques while preserving physical consistency, *Journal Soft Computing*, 25:5723–5746.
- Norouzi, H., Nadiri, A.A., Asghari Mogaddam, A., & Gharekhani, M. (2017). Prediction of Transmissivity of Malikan Plain Aquifer Using Random Forest Method. *Water and Soil Science*, 27(2), 61-75 (In Persian).
- Nou, M. R. G., Foroudi, A., Latif, S. D., & Parsaie, A. (2022). Prognostication of scour around twin and three piers using efficient outlier robust extreme learning machine. *Environmental Science and Pollution Research*, 29(49), 74526-74539.
- Pandey, M., Ahmad, Z., & Sharma, P. K. (2016). Estimation of maximum scour depth near a spur dike. *Canadian Journal of Civil Engineering*, 43(3), 270-278.
- Pandey, M., Jamei, M., Ahmadianfar, I., Karbasi, M., Lodhi, A.S., & Chu, X. (2022). Assessment of scouring around spur dike in cohesive sediment mixtures: A comparative study on three rigorous machine learning models. *Journal of Hydrology*, 606, 127330.
- Parisouj, P., Mohebzadeh, H., & Lee, T. (2020). Employing machine learning algorithms for streamflow prediction: a case study of four river basins with different climatic zones in the United States. *Water Resources Management*, 34, 4113-4131.
- Parsaie, A., Haghiabi, A. H., & Moradinejad, A. (2019). Prediction of scour depth below river pipeline using support vector machine. *KSCE Journal of Civil Engineering*, 23, 2503-2513.
- Roushangar, K., Goodarzi, S., & Abbaszadeh, H. (2024). Numerical Investigation of the Performance of Blade Groynes on Scouring and its Effect on Hydraulic Parameters of Sediment and Flow. *Environment and Water Engineering*, 10(1), 121-136 (In Persian).
- Shojaeian, Z., Kashfipour, S.M., Mosavi Jahromi, S.H., & Shafae Bajestan, M. (2014). Experimental Study on The Local Scouring of Series of Bandal-Like spurs in Clear Water Condition. *Journal of Irrigation Sciences and Engineering*, 38(2), 21-32 (In Persian).
- Teraguchi, H., Nakagawa, H., Kawaike, K., Bans, Y., & Zhang, H. (2011). Effects of hydraulic structures on river morphological processes. *International Journal of Sediment*, 26(3), 283-303.
- Tikhmarine, Y., Souag-Gamane, D., & Kisi, O. (2019). A new intelligent method for monthly streamflow prediction: hybrid wavelet support vector regression based on grey wolf optimizer (WSVR–GWO). *Arabian Journal of Geosciences*, 12, 1-20.
- Tripathi, R. P., & Pandey, K. (2023). Gene-expression programming for scour around spur dike. *International Journal of Hydrology Science and Technology*, 15(3), 295-303.
- Wang K., Wen, X., Hou, D., Tu, D., Zhu, N., Huang, P., & Zhang, H. (2018). Application of least-squares support vector machines for quantitative evaluation of known contaminant in water distribution system using online water quality parameters. *Sensors*, 18(4), 938.

# Evaluation of the performance of machine learning methods for estimating the maximum scour depth around the bandallike spur-dike

## EXTENDED ABSTRACT

### Introduction

Various methods have been proposed to control riverbank erosion, with one of the most common methods being the use of a spur-dike. One of the key issues with these structures is the scouring around their foundations due to changes in flow patterns. Scouring around the foundation can lead to structural instability and ultimately its destruction. To reduce the depth of scouring around the foundation of the dam structure, a new type of spur-dike "BandalLike" has been introduced. This structure consists of a permeable part at the bottom and an impermeable part on top. Experimental relationships have been established to estimate the maximum depth of scouring around the foundation of the impermeable spur-dike to consider precautions in design. However, there are not many relationships provided for the permeable dam structure. Therefore, the aim of the current research is to evaluate the performance of machine learning-based methods to estimate the maximum depth of scouring around the "BandalLike" structure.

### Methodology

In the current research, the results of a laboratory experiment were used. Dimensionless influential variables (input variables) on the scouring around the bandallike were considered, including Froude number ( $Fr$ ), angle of installation of the bandallike relative to the flume wall ( $\theta$ ), percentage of permeability of the porous section ( $\alpha$ ), and the distance ratio to the length of the bandallike ( $S/L$ ). It is worth mentioning that these variables were determined after dimensional analysis. Additionally, the ratio of maximum scouring depth to the flow depth ( $h_s/h$ ) was considered as the output variable. For each variable, a total of 108 data series were extracted. 80% of the data series (86 series) were used for model training, and the remaining 20% (22 series) were used for model evaluation. Furthermore, GEP, SVM, and RF methods were used in this research.

### Results and Discussion

For all models, the best performance for single-variable scenarios was determined to be scenario S2 with input parameter  $\alpha$ . In this regard, the MAE index for GEP, SVM, and RF models was estimated to be 0.077, 0.067, and 0.073, respectively. Additionally, the lowest performance level was determined for scenario S3 with input parameter  $S/L$ . For two-variable scenarios, the results showed that the best input combination for all models is scenario S5 with variables  $\alpha$  and  $Fr$ . Based on this, the RMSE index for GEP, SVM, and RF models was determined to be 0.015, 0.015, and 0.009, respectively. With an increase in the number of inputs to two variables, the average MAE index for GEP and RF models decreased by 12% and 24%, respectively, indicating increased accuracy with an increase in inputs for these two models. However, the average MAE index for the SVM model increased by 12%, indicating a decrease in evaluation accuracy for this model with an increase in inputs. For three-variable scenarios, the GEP model achieved the highest accuracy for scenario S14 with parameters  $\alpha$ ,  $Fr$ , and  $\theta$ , with RMSE, MAE, and CC statistical indices of 0.013, 0.017, and 0.990,

respectively. In SVM and RF models, the best three-variable scenario was S11 with parameters  $\alpha$ , Fr, S/L, with RMSE evaluation indices of 0.009 and 0.008, respectively. It was found that the RF model, like the GEP model, increased accuracy (a 41% decrease in average MAE index) with an increase in the number of input variables. However, increasing the number of inputs did not significantly change the accuracy of the SVM model. For four-variable scenarios (S15), the SVM model faced a significant decrease in accuracy, with an approximate 140% increase in the average MAE index. Meanwhile, the GEP and RF models also experienced increased accuracy for scenario S15.

### **Conclusion**

All methods in single-variable input mode had the most and least impact respectively related to parameters  $\alpha$  and S/L. In the SVM model, increasing the number of inputs from single-variable to two-variable mode resulted in an almost 2-fold increase in the average MAE index. In the GEP model, increasing the number of inputs from three variables to four variables led to an approximately 3.5-fold increase in the average MAE index. However, in the RF method, increasing the number of inputs resulted in improved model accuracy, with the average MAE index in the four-variable mode decreasing by 83% compared to the three-variable mode. Ultimately, it was evident that the RF method provided much better performance in estimating the depth of sedimentation around the bandallike spur-dike compared to other methods.

**Keywords:** Scouring, Artificial intelligence, River, spur-dike.