ارزیابی ضریب دبی جریان سرریزهای با محیط متخلخل با استفاده از آزمایش‌های هیدرولیکی و مدل‌های داده مبنا

سید علیرضا اسماعیلی[۱]، مجید حیدری[۲*] و سعید گوهری[۳]

۱ -دانشجوی دکتری علوم ومهندسی آب- سازه های آبی، گروه علوم ومهندسی آب، دانشکده کشاورزی دانشگاه بوعلی سینا.

۲*- نویسنده مسئول، دانشیار گروه علوم ومهندسی آب، دانشکده کشاورزی دانشگاه بوعلی سینا mheydari@basu.ac.ir

۳- دانشیار گروه علوم ومهندسی آب، دانشکده کشاورزی دانشگاه بوعلی سینا.

دریافت: بازنگری: پذیرش: انتشار:

چکیده

سرریزهای گابیونی لبه‌پهن مستطیلی به‌دلیل قابلیت بالای کنترل جریان، کاهش انرژی و سازگاری زیست‌محیطی، کاربرد گسترده‌ای در سازه‌های هیدرولیکی دارند. در این پژوهش، ضریب دبی جریان عبوری از این نوع سرریزها بر اساس داده‌های آزمایشگاهی و با استفاده از مدل‌های یادگیری ماشین پیش‌بینی شد. بدین منظور، پنج مدل یادگیری ماشین مبتنی بر رگرسیون نظیر RF، SR، SVR، XGBoost و CatBoost توسعه داده شد. نتایج نشان داد که مدل‌های CatBoost و XGBoost بالاترین دقت را در پیش‌بینی مقادیر ضریب سرریز ارائه می‌دهند. تحلیل کلی عملکرد مدل‌ها نشان داد که تقسیم‌بندی نظیر ۸۰ درصد داده‌های خام برای آموزش مدل نتایج بسیار خوبی به همراه داشته است. به‌طوری که مدل‌های XGBoost و CatBoost با R^2 به‌ترتیب برابر با ۰/۹۹۰ و ۰/۹۹۶ ، و RMSE معادل با ۰/۰۳۸و ۰/۰۱۹ بهترین نتایج را به همراه داشته است. اما مدل‌های RF، SR و SVR در تمامی سناریوها ضعیف‌ترین عملکرد را داشتند. همچنین استخراج رابطه تجربی با استفاده از الگوریتم GEP نشان داد که نسبت آموزش و آزمون مدل نقش تعیین‌کننده‌ای در دقت معادله نهایی دارد.

کلید واژه‌ها: سرریز گابیونی، جریان آزاد، یادگیری ماشین، XGboost، CatBoost، GEP.

مقدمه

رفتار هیدرولیکی سرریزهای گابیونی به‌واسطه ماهیت متخلخل آن‌ها، تفاوت اساسی با سرریزهای صلب دارد. برخلاف سرریزهای معمولی، در این سازه‌ها جریان به دو بخش مجزا تقسیم می‌شود: جریان از روی تاج و جریان نفوذی از داخل بدنه متخلخل سازه عبور می‌کند. این ویژگی، در کنار زبری ذاتی سطح گابیون، موجب افزایش چشمگیر آشفتگی جریان و استهلاک مؤثر انرژی جنبشی می‌شود. در طراحی این سازه‌ها، ضریب دبی کلیدی‌ترین پارامتر است که مستقیماً بر ظرفیت عبور جریان تأثیر می‌گذارد. نفوذ جریان به داخل بدنه سرریز، علاوه بر تأثیر بر پروفیل سطح آب و توزیع فشار زیر سازه، منجر به کاهش ضریب تخلیه در مقایسه با سرریزهای صلب با هندسه مشابه می‌گردد (Esmaili et.al, 2026). با این حال، پیش‌بینی دقیق ضریب دبی این سازه‌ها به‌دلیل ماهیت پیچیده جریان متخلخل و وابستگی آن به پارامترهای متعدد هندسی و هیدرولیکی نظیر ارتفاع تاج، شیب پایین‌دست، دانه‌بندی مصالح و شرایط جریان بالادست بسیار دشوار است. مطالعه‌های پیشین عمدتاً بر شناخت رفتار فیزیکی و مزایای اکولوژیک این سازه‌ها متمرکز بوده‌اند. به‌عنوان نمونه، پژوهش Kumar et. al (2025) نشان داد که سرریزهای گابیونی در رودخانه‌های دارای بستر دگرگون‌شده، نقش به‌سزایی در تسهیل عبور ماهیان و بهبود تنوع زیستی دارند. Pallavi و Harashit (2018) ویژگی‌های جریان عبوری از درون و روی سرریزهای گابیونی را به‌صورت تجربی مطالعه کردند. دو مدل سرریز از نوع گابیونی و صلب در کانالی به طول ۴/۷ متر، عرض ۰/۲۱ متر و ارتفاع ۰/۳ متر استفاده بررسی کردند. ابعاد مدل‌ها سرریز برای نوع گابیونی و صلب یکسان بود با مشخصات ارتفاع ۰/۱۰۵ متر، عرض ۰/۲۱ متر و طول ۰/۱ متر و دو نوع دانه‌بندی با متوسط ۱۶ و ۲۰ میلی‌متر به‌عنوان مصالح سنگ‌ریزه برای بررسی تأثیر اندازه سنگ‌دانه‌ها بر کاهش سرعت استفاده شد. نتایج نشان داد که سرریزهای گابیونی در کاهش سرعت جریان نسبت به سرریزهای صلب مؤثرتر بودند. علاوه بر این، با افزایش قطر مصالح، هد آب بالای سرریز گابیونی کاهش یافت. Jiang et.al (2018) به بررسی اثر زاویه بالادست بر رفتار جریان، ضریب دبی و ناحیه جدایش در سرریزهای تخت ذوزنقه‌ای با استفاده از نرم‌افزار OpenFOAM پرداختند. یافته‌ها حاکی از آن بود که زاویه بالادست تأثیر مستقیم بر ضریب دبی و الگوی جدایش دارد و تغییر آن موجب تغییر قابل‌توجه رفتار جریان می‌شود. هر دو مدل آشفتگی k–ε و k–ω SST نتایج رضایت بخشی ارائه کردند به‌طوری که مدل k–ω SST اندکی برتری و خطای کمتر نشان داد. Shariq et.al (2022) در پژوهشی بر روی سرریزهای گابیونی و صلب تحت شرایط جریان آزاد و جریان مستغرق مطالعه‌های آزمایشگاهی انجام دادند. آن‌ها پروفیل سطح آب بر روی سرریز صلب و گابیونی با تخلخل‌های مختلف بررسی کردند و نتیجه گرفتند که نسبت افت انرژی به ارتفاع تاج سرریز گابیونی بر ضریب دبی سرریز تأثیرگذار است. Biabani et.al (2022)، اثر شیب

بالادست و پایین‌دست و ویژگی‌های تخلخل بر نحوه جریان، ضریب دبی و اتلاف انرژی در سرریزهای گابیونی بررسی کردند. نتایج نشان داد که با کاهش شیب بالادست، ضریب دبی کاهش می‌یابد، اما سرریزهای گابیونی در همه حالت‌ها ضریب دبی بیشتری نسبت به سرریزهای صلب دارند. همچنین با افزایش دبی ورودی، اتلاف انرژی تا حدود ۴۵ درصد کاهش پیدا کرد. Doustkam et.al (2024)، عملکرد هیدرولیکی سرریزهای متخلخل تخت با تاج شیب‌دار در شرایط جریان آزاد و مستغرق ارزیابی کردند. نتایج آزمایشگاهی نشان داد که اعمال شیب در تاج سرریز باعث افزایش قابل‌توجه ضریب دبی می‌شود؛ به‌طوری‌که شیب بالادست ۳۸ درصد و شیب پایین‌دست ۲۸ درصد نسبت به نوع صلب عملکرد بهتری داشت. همچنین سرریزهای با تاج شیب‌دار نسبت به تغییرات سطح آب پایین‌دست حساس‌تر بودند و روابط تجربی دقیق‌تری برای پیش‌بینی ضریب دبی و کاهش دبی در جریان مستغرق ارائه شد. Alomari et.al (2023)، به بررسی تخمین ضریب دبی با استفاده از عمق روی لبه سرریزهای گابیونی لبه پهن مستطیلی پرداخته است. نتایج نشان‌دهنده افزایش ۱۰/۵ درصدی در ضریب دبی نسبت به سرریزهای صلب است. Haghiabi et.al (2018)، ضریب دبی سرریز مثلثی با استفاده از شبکه عصبی پرسپترون چندلایه (MLP) و سیستم استنتاج فازی عصبی تطبیقی (ANFIS) مورد بررسی قرار دادند. نتایج مدل ANFIS با شاخص‌های خطا شامل ضریب تبیین ۰/۹۷ و ریشه میانگین مربعات خطا ۰/۰۳ بسیار مناسب بود. مقایسه نتایج مدل MLP با مدل ANFIS نشان داد که هر دو مدل عملکرد بسیار مناسبی دارند، بااین‌حال ساختار مدل ANFIS بهینه‌تر است. Akbari et.al (2019)، کاربرد محاسبات نرم در تخمین ضریب سرریز کلید پیانویی با تاج بلند با استفاده از روش‌های MLP، GPR، SVM، GRNN، رگرسیون‌های خطی چندگانه و غیرخطی بررسی کردند. نتایج این مطالعه نشان داد که روش GPR به‌عنوان یک روش جدید و مؤثر برای پیش‌بینی ضریب دبی سرریزها با RMSE برابر ۰/۰۱۱، R^2 برابر ۰/۹۹۲ معرفی شد و در مقایسه با سایر روش‌ها بهترین نتایج را ارائه داد. در سال‌های اخیر، استفاده از مدل‌های یادگیری ماشین برای پیش‌بینی عملکرد هیدرولیکی سرریزهای گابیونی نیز افزایش‌یافته است. مدل‌های مبتنی بر SVM به‌عنوان روشی قوی برای مدیریت داده‌های غیرخطی شناخته می‌شوند. این مدل‌ها در پیش‌بینی رواناب و دبی رودخانه‌ها کاربرد داشته‌اند، اما کاربرد مستقیم آن‌ها در سرریزهای گابیونی محدود بوده است. مدل‌هایی مانند CatBoost و XGBoost قادر به پیش‌بینی دقیق ضریب دبی و نسبت دبی عبوری هستند و در شرایط جریان آزاد عملکرد بهتری نسبت به روش‌های سنتی ارائه می‌دهند. Seif et.al (2024)، با استفاده که روش‌های فرا ابتکاری پیش‌بینی ضریب دبی جریان مستغرق برای سرریزهای لبه پهن متخلخل (PBCWs) بررسی کردند. عملکرد مدل ترکیبی MLP-GNDO با مدل‌های MLP، برنامه‌ریزی بیان ژن (GEP) و رگرسیون غیرخطی استاندارد (SNR) مقایسه شد. نتایج نشان داد که مدل ترکیبی MLP-GNDO با دستیابی به خطای جذر میانگین مربعات ۰/۰۲۱ و ۰/۰۲۲ و مقدار R^2 برابر با ۰/۹۶۴ و ۰/۹۵۴ به‌ترتیب برای مجموعه داده‌های آموزشی و آزمایشی، از مدل‌های MLP، GEP و SR عملکرد بهتری داشته است. Rahmanshahi et al (2024)، به بررسی دقیق عملکرد هیدرولیکی سرریزهای خاکریز متخلخل (PEWs) پرداخته و به‌دلیل پیچیدگی جریان در این سازه‌ها، ترکیب شبکه عصبی پرسپترون چندلایه (MLP) با پنج الگوریتم متاهیورستیک جدید شامل GNDO، RUN، SMA، WOA و GWO را برای پیش‌بینی سه شاخص کلیدی یعنی ضریب دبی آزاد (FDC)، شاخص حد مدولار (MLI) و ضریب کاهش دبی در شرایط غرقاب (SDRF) معرفی کرده است. نتایج نشان داد که مدل GNDO-MLP بهترین عملکرد را داشته و نسبت به سایر الگوریتم‌ها و حتی روش کلاسیک Levenberg-Marquardt بهبود چشمگیری در دقت و کاهش عدم قطعیت ارائه کرده است، به‌گونه‌ای که توانست روابط غیرخطی پیچیده جریان در سازه‌های متخلخل را باقابلیت اعتماد بالا مدل‌سازی کند.

Nourani et.al (2021)، در پژوهشی با هدف تخمین دقیق ضریب دبی در سرریزهای لبه پهن غیرمتخلخل، ابتدا داده‌های عددی حاصل از شبیه‌سازی جریان با نرم‌افزار ANSYS FLUENT و روش حجم محدود (FVM) را تولید کرده و سپس دو مدل هوش مصنوعی شبکه عصبی مصنوعی (ANN) و رگرسیون فرایند گاوسی (GPR) را توسعه داده است. برای ارتقای عملکرد این مدل‌ها، الگوریتم (HHO) به‌کار گرفته شد و مدل‌های ترکیبی ANN-HHO وGPR-HHO معرفی شدند. نتایج نشان دادند که مدل ANN-HHO نسبت به سایر مدل‌ها دقت بالاتری در پیش‌بینی ضریب دبی دارد و توانسته است روابط غیرخطی پیچیده بین پارامترهای هندسی مانند (h/L_r، h/P، شیب‌های بالادست و پایین‌دست و شیب تاج) و ضریب دبی را با خطای کمتر مدل‌سازی کند.

Azma et.al (2023)، به بررسی عملکرد هیدرولیکی سرریزهای لبه پهن گابیونی پرداخته است که به‌دلیل تخلخل، جریان علاوه بر عبور از تاج، از داخل بدنه نیز عبور می‌کند. داده‌های آزمایشگاهی موجود از مطالعه‌های پیشین برای آموزش مدل‌ها استفاده شد و الگوریتم‌های یادگیری ماشین و ensemble boosting شامل XGBoost، CatBoost و Gradient Boosting با مدل‌های پایه مانند RF و SR مقایسه شدند. نتایج نشان دادند که مدل CatBoost بهترین عملکرد را در تخمین نسبت جریان عبوری از بدنه (Q_{in}/Q_t) با ضریب تبیین ۰/۹۸۲ داشته و در تخمین ضریب دبی نیز مدل‌های CatBoost و XGBoost به‌ترتیب ۰/۹۹۴ و ۰/۹۹۲ بسیار دقیق عمل کردند، درحالی‌که مدل‌های ساده مانند درخت تصمیم ضعیف‌ترین نتایج را ارائه دادند.

Nourani et al (2023)، از مدل‌های یادگیری ماشین نظارت‌شده برای پیش‌بینی ضریب دبی در سرریزهای گابیونی لبه پهن مستطیلی استفاده کرده است. نتایج نشان‌دهنده دقت بالای مدل‌های SVM و شبکه عصبی در پیش‌بینی ضریب دبی است. Tiwari et al (2023)، مدل‌های مختلف یادگیری ماشین برای پیش‌بینی عملکرد هوادهی سرریزهای گابیونی لبه پهن مستطیلی موردبررسی قراردادند. نتایج نشان‌دهنده عملکرد برتر شبکه‌های عصبی عمیق در پیش‌بینی کارایی هوادهی است.

Hameed et al (2025)، به استفاده از مدل‌های ترکیبی تقویتی برای تخمین ضریب دبی و دبی عبوری در سرریزهای گابیونی لبه پهن مستطیلی پرداخته است. نتایج نشان‌دهنده دقت بالای مدل‌های ترکیبی در پیش‌بینی پارامترهای هیدرولیکی است.

Shakeri yousefi et al (2025)، به بررسی پیش‌بینی دقیق اتلاف انرژی در سیل بندهای گابیونی پایین‌دست سرریزها پرداخته است. سه مدل یادگیری ماشین شامل (SVR)، (GEP) و (ANN) توسعه و مقایسه شدند. نتایج نشان دادند که چهار پارامتر کلیدی عمق اولیه پرش هیدرولیکی، عمق متوالی پرش، عمق جریان پس از گابیون و عدد فرود اولیه بیشترین تأثیر را بر اتلاف انرژی دارند. در ارزیابی عملکرد، مدل GEP بهترین نتایج را ارائه کرد و با ضریب تبیین ۰/۹۳۶ و خطای بسیار پایین ۰/۰۰۳ توانست از ANN و SVR به‌طور قابل‌توجهی بهتر عمل کند.

Khatibi et.al (2014)، در مطالعه آزمایشگاهی اتلاف انرژی در سرریزهای گابیونی پله‌ای، عملکرد برنامه‌نویسی بیان ژن (GEP)را در کنار شبکه‌های عصبی مصنوعی (ANN) بررسی کرده و نشان دادند که این روش قادر به بازنمایی دقیق روابط هیدرولیکی پیچیده میان متغیرهای جریان است.

همچنین سایر پژوهشگران ازجمله، kim et al (2014)، Salmasi و Abraham (2020)، Salmasi et al (2021)، Rahmanshahi و Shafai Bejestan (2020)، Roushangar et al (2010)، Azamathulla et al (2014)، عملکرد مناسب برنامه‌نویسی بیان ژن در پژوهش‌های آزمایشگاهی از جمله برآورد ضریب دبی در سرریز مستطیلی لبه‌پهن، ضریب سرریزهای اوجی، ضریب دبی سرریزهای گابیونی قوسی، افت انرژی سرریزهای پلکانی، افت انرژی در بلوک‌های شیب‌دار، ضریب زبری، فرسایش بستر، پیش‌بینی سرعت در کانال‌های مرکب، را تأیید کردند.

تاکنون، مدل‌های مختلف یادگیری ماشین به‌عنوان مدل‌های جانشین برای برآورد ضریب دبی انواع مختلف سرریزهای صلب مورد استفاده قرار گرفته‌اند. با این حال، تنها تعداد محدودی از مطالعه‌های بر سرریزهای گابیونی تمرکز داشته‌اند و هیچ مطالعه‌ای جریان عبوری در سرریزهای متخلخل را مورد بررسی قرار نداده است. این امر ناشی از مشکلات عملی مرتبط با محاسبه و اندازه‌گیری این نوع جریان‌هاست که در نهایت به کمبود یک مجموعه داده‌ی قابل‌اعتماد برای چنین اندازه‌گیری‌هایی منجر شده است. هدف از این پژوهش استفاده از مدل‌های داده مبنا به‌منظور برآورد ضریب دبی گذرنده از سرریزهای گابیونی لبه‌پهن مستطیلی، مقایسه نتایج حاصل از این مدل‌ها با داده‌های آزمایشگاهی انجام شده و در نهایت استخراج رابطه برای پارامتر مزبور با استفاده از الگوریتم بیان ژن (GEP) می‌باشد.

به‌رغم انجام مطالعه‌های گسترده آزمایشگاهی و عددی، توسعه مدل‌های دقیق برای تخمین ضریب دبی این سازه‌ها همچنان با محدودیت روبه‌رو است. اخیراً کاربرد مدل‌های یادگیری ماشین به دلیل توانایی بالا در بازنمایی روابط غیرخطی و پیچیده، در مهندسی هیدرولیک توسعه یافته است. با این حال، بررسی مقایسه‌ای عملکرد مدل‌های نوین و پیشرفته مبتنی بر یادگیری تجمعی در کنار استخراج روابط تحلیلی صریح برای سرریزهای با محیط متخلخل، همچنان در ادبیات فن محدود است. بر این اساس، شکاف تحقیقاتی حاضر، عدم وجود یک چارچوب جامع داده‌مبنا برای پیش‌بینی دقیق ضریب دبی با در نظر گرفتن اثر توأمان تخلخل و شرایط هیدرودینامیکی است.

در راستای پر کردن این خلأ علمی، نوآوری و وجه تمایز پژوهش حاضر در ترکیب داده‌های حاصل از مطالعه‌های آزمایشگاهی با مدل‌های نوین داده‌مبنا است تا علاوه بر ارزیابی جامع رفتار جریان در سرریزهای متخلخل، یک رابطه ریاضی صریح، نوین و با دقت بالا برای برآورد ضریب دبی ارائه گردد که قابلیت استفاده مستقیم در طراحی‌های مهندسی را داشته باشد.

مواد و روش‌ها

تحلیل ابعادی

روابط حاکم بر بسیاری از پدیده‌های هیدرولیکی را نمی‌توان مستقیماً از قوانین عمومی سیالات یا تئوری‌های موجود، مانند قوانین بقای جرم، انرژی و ممنتوم، استخراج کرد. برای تحلیل هیدرولیکی جریان در سرریزهای گابیونی، ابتدا تلاش شد با استفاده از تحلیل ابعادی و بر اساس تئوری باکینگهام، پارامترهای بدون‌بعدی برای محاسبه ضریب سرریز تعیین شوند. چندین رابطه برای محاسبه دبی جریان در سرریزهای لبه پهن مستطیلی ارائه شده است. شکل کلی رابطه دبی جریان روی یک سرریز لبه پهن به‌صورت معادله (۱) می‌باشد.

$$Q_w = \frac{2}{3} C_d B \sqrt{\frac{2}{3} g {H_o}^3} \qquad (۱)$$

$$H_o = H + \left(\frac{1}{2g}\right)\left(\frac{Q}{B(H+P)}\right)^2 \qquad (۲)$$

در رابطه (۱) و (۲)، (H) اختلاف ارتفاع بین سطح آب بالادست و تاج سرریز که در فاصله 4H، (H_o) انرژی مخصوص و (P) ارتفاع سرریز است. باتوجه به شکل (۱)، متغیرهای مؤثر بر ضریب سرریز (C_d) در شرایط جریان آزاد به عمق آب بالادست (D)، طول سرریز (L)، قطر متوسط مصالح سنگی (d_{50})، عرض

کانال اصلی(B)، سرعت جریان در بالادست(V_u)، تخلخل ذرات(φ)، شتاب ثقل(g)، جرم واحد حجم مایع(ρ) لزجت دینامیکی(μ) است که به صورت یک رابطه تابعی زیر می‌توان نوشت:

$$C_d = f(B, H, P, L, V_u, D, \rho, \mu, g, d_{50}, \varphi) \qquad (۳)$$

با در نظر گرفتن (ρ)، (V_u) و (L) به عنوان متغیرهای تکراری، رابطه تابعی برای (C_d) بر حسب پارامترهای بدون بعد را می‌توان به‌صورت زیر نوشت:

$$C_d = f(\frac{\mu}{\rho V_u L}, \frac{H}{P}, \frac{H}{L}, \frac{Dg}{V_u^2}, \frac{d_{50}}{P}, \varphi) \qquad (۴)$$

$$C_d = f(\frac{H}{L}, \frac{H}{P}, \frac{d_{50}}{P}, Fr = \frac{V_u}{\sqrt{gD}}, Re = \frac{\rho Q_w}{B\mu}, \varphi) \qquad (۵)$$

با توجه به اینکه در شرایط جریان آزاد به‌دلیل آشفتگی جریان اثر نیروی لزجت بر پارامترهای هیدرولیکی ناچیز است لذا می‌توان از عدد رینولدز (Re) آن صرف‌نظر کرد. بنابراین می‌توان نتیجه گرفت که ضریب دبی سرریز گابیونی در شرایط جریان آزاد به‌صورت تابعی از پارامترهای بدون بعد زیر است. بدین ترتیب، چارچوب نظری لازم برای انتخاب متغیرهای ورودی مدل‌های یادگیری ماشین فراهم گردید.

$$C_d = f(\frac{H}{L}, \frac{H}{P}, Fr, \frac{d_{50}}{P}, \varphi) \qquad (۶)$$

تجهیزات آزمایشگاهی

آزمایش‌ها در یک کانال آزمایشگاهی به طول هفت متر، عرض ۵۰ سانتی‌متر و ارتفاع ۶۰ سانتی‌متر در آزمایشگاه هیدرولیک گروه علوم ومهندسی آب دانشگاه فردوسی مشهد انجام شده است (شکل ۲). دیواره‌های فلوم به منظور امکان روئیت جریان، شیشه‌ای و شیب فلوم ثابت و برابر ۰/۰۰۱ است. آبراهه مجهز به دبی‌سنج اولتراسونیک برای اندازه‌گیری جریان و سیستم اینورتر برای تنظیم جریان می‌باشد. سرریز مستطیلی هم‌عرض آبراهه و مجموعه شبکه آرام‌کننده ورودی جریان، و نیزدریچه کشویی برای کنترل عمق آب پایاب در انتها می باشد. حداکثر دبی تولیدی توسط پمپ ۶۰ لیتر بر ثانیه است. اندازه‌گیری پروفیل سطح آب از عمق سنج مکانیکی با ±۰/۱ میلی‌متر استفاده شد. دراین پژوهش هشت مدل سرریز گابیونی بررسی شد. آزمایش‌ها با محدوده‌های مختلف اندازه متوسط سنگ‌دانه‌ها، طول سرریز و دبی انجام شد. همچنین برای تعیین اندازه متوسط سنگ‌دانه‌ها از روش استاندارد ASTM استفاده گردید.

شکل (۳) منحنی دانه‌بندی سنگدانه‌ها مورد آزمایش را نشان می‌دهد. ضریب یکنواختی دانه‌بندی سنگدانه‌ها برابر ۱/۳۱ که با توجه به معیار یکنواختی دانه‌بندی در حدود یک تا ۱/۳ بعنوان مصالح یکنواخت در نظر گرفته شد (Heravi et.al, 2024).

از این رو می‌توان با اطمینان از تأثیر انحراف معیار هندسی توزیع اندازه ذرات گابیون صرف نظر کرد. در جدول (۱) مدل سریزهای مورد استفاده در آزمایش و پارامترهای هیدرولیکی ارائه شده است.

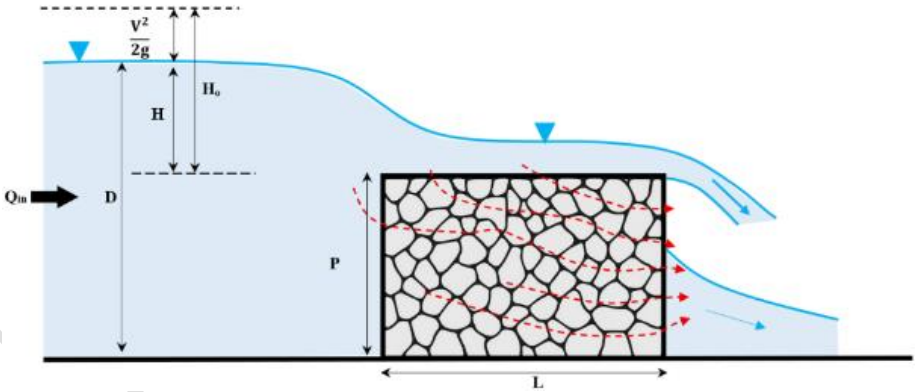


Fig. 1- Hydraulic parameters in a rectangular Broad-Crested gabion weir

شکل ۱- پارامترهای هیدرولیکی در سرریز گابیونی لبه‌پهن مستطیلی

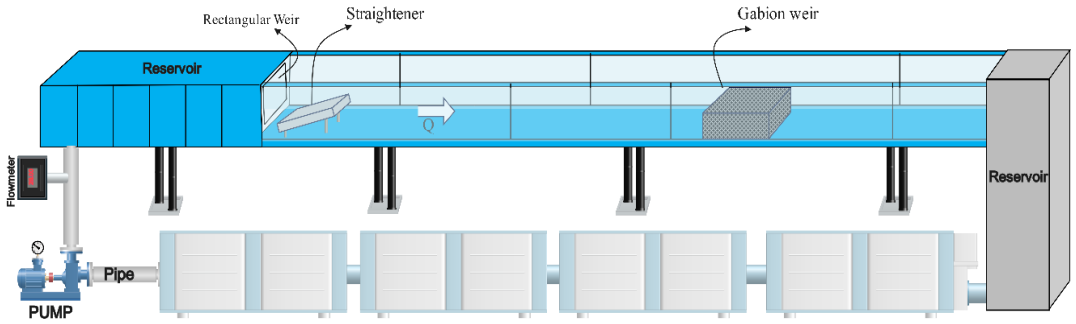


Fig. 2- Laboratory channel and speed measuring instrument

شکل۲- کانال آزمایشگاهی و ابزار اندازه‌گیری سرعت

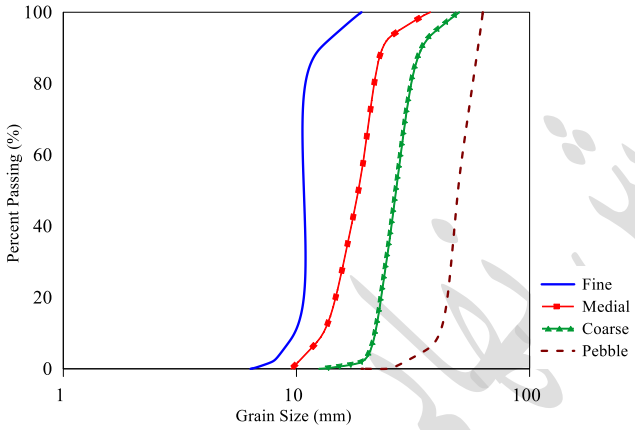


Fig. 3 - The used grading curve

شکل۳- منحنی دانه‌بندی مورد استفاده

جدول ۱- مدل‌های سرریز محیط متخلخل مورد استفاده

Table 1- Porous media overflow models used

Parameters	Dimension	Range
Intake discharge, Q	L/s	10-60
Length, L	mm	300,600
hight , P	mm	200
Grain size, d_{50}	mm	10.8, 18.5, 28.4, 37.8
Porousity, φ	dimensionless	28.6, 34.5, 36, 41

مدل‌های داده مبنا استفاده شده

کاربرد مدل‌های تجمیعی (Ensemble) و به‌ویژه مدل‌های جدیدتر و پیشرفته‌تری مانند XGBoost و CatBoost نیز به تعداد محدودی از مطالعه‌ها محدود بوده است. از این‌رو، در این مقاله از برخی مدل‌های پایه شامل رگرسیون بردار پشتیبان (SVR) به‌همراه چندین مدل تجمیعی استفاده شده است؛ از جمله جنگل تصادفی (RF) به‌عنوان یک مدل تجمیعی مبتنی بر bagging، مدل رگرسیون تجمیعی (SR)، و دو مدل مختلف مبتنی بر تقویت Boosting شامل مدل XGBoost، و مدل CatBoost می‌باشد که در ادامه مدل‌های مورد استفاده تشریح شده‌اند. برای مدل‌سازی از زبان برنامه نویسی پایتون برای مدل‌های مبتنی بر درخت‌های رگرسیونی به کمک کتابخانه‌های اصلی شامل pandas و numpy نسخه ۲.۳.۱ و کتابخانه‌های یادگیری ماشین از جمله scikit-learn نسخه ۱.۸.۰، RandomForestRegressor، sklearn.svm، CatBoostRegressor نسخه ۱.۲.۸ و زبان برنامه‌نویسی متلب نسخه ۲۰۱۸ برای مدل‌های SVR و XGBoost استفاده شد.

مدل Random Forest (RF)

مدل جنگل تصادفی یک روش گروهی است که مجموعه‌ای از درخت‌های رگرسیونی را با استفاده از نمونه‌برداری Bootstrap و انتخاب تصادفی ویژگی‌ها تولید می‌کند. این روش با کاهش واریانس و افزایش پایداری، تخمینی مقاوم برای مسائل رگرسیونی غیرخطی فراهم می‌کند. این امر با اختصاص نمونه‌های داده و

زیرمجموعه‌های ویژگی مختلف از مجموعه آموزشی و همچنین تقسیم زیرمجموعه‌ها در درخت تصمیم‌گیری بر اساس ویژگی‌های مختلف حاصل می‌شود. برای آموزش درخت‌های یادگیری T_b، از داده‌های آموزشی اصلی برای تولید نمونه‌های بوت‌استرپ استفاده می‌شود. می‌توان نمونه‌های بوت‌استرپ را با نمونه‌های جدید جایگزین کرد. با استفاده از بردار داده‌های ورودی، x و B درخت رگرسیون متمایز برای هر نمونه بوت‌استرپ ساخته می‌شوند. پیش‌بینی‌های جنگل تصادفی با میانگین‌گیری از پیش‌بینی‌های B درخت رگرسیون، $T_b(x)$ تولید می‌شوند. هر درخت با استفاده از یک زیرمجموعه تصادفی از داده‌ها و یک زیرمجموعه ویژگی در هر گره ساخته می‌شود. معیار اصلی تقسیم گره‌ها کاهش واریانس است:

$$\Delta V = Var(t) - \left(\frac{N_L}{N} Var(t_L) + \frac{N_R}{N} Var(t_R)\right) \quad (۷)$$

در این رابطه، t: گره والد، t_L, t_R برابر گره‌های فرزند چپ و راست، N_L, N_R تعداد نمونه‌ها در گره‌های چپ و راست، $\mathrm{Var}(t_R)$ و $\mathrm{Var}(t_L)$ واریانس گره‌های فرزند و ΔV مقدار بهبود حاصل از تقسیم گره می باشند که در حالت مقدار بشینه گره انتخاب خواهد شد. خروجی هر درخت برای ورودی x برابر میانگین مقادیر هدف در برگ مربوطه است:

$$T_b(x) = \overline{y}_{leaf} \quad (۸)$$

$\overline{y}_{leaf}$ میانگین مقادیر هدف نمونه‌های قرارگرفته در برگ مربوط به و پیش‌بینی نهایی مدل برابر میانگین خروجی تمام درخت‌هاست:

$$\widehat{y} = \frac{1}{B}\sum_{b=1}^{B} T_b(x) \quad (۹)$$

مدل Stacked Regression (SR)

رگرسیون انباشته یک چارچوب یادگیری تجمیعی است که در آن چند مدل پایه به‌طور مستقل روی داده اصلی آموزش داده می‌شوند و خروجی آن‌ها به‌عنوان ورودی یک متا-مدل سطح بالا مورد استفاده قرار می‌گیرد (Roustaei, 2024). برای جلوگیری از بیش‌برازش، خروجی مدل‌های پایه از طریق پیش‌بینی‌های برون‌تایی و با استفاده از K-Fold Cross-Validation تولید می‌شود. برای هر نمونه $\mathbf{x}_i$:

$$z^{O}_{(i,m)} = f_m^{(-k)}(x_i), x_i \in Fold_k \quad (۱۰)$$

در این رابطه $z^{O}_{i,m}$ خروجی مدل پایه m برای نمونه i که مدل روی داده خارج از Fold آموزش دیده، $f_m^{(-k)}$ مدل پایه m که روی همه Fold ها به‌جز Fold شماره k آموزش دیده، Fold_k بخش k-اُم داده در Cross-Validation و K تعداد Fold ها است. ماتریس حاصل از این پیش‌بینی‌ها مجموعه ویژگی‌های جدیدی ایجاد می‌کند که متامدل روی آن آموزش می‌بینند.

$$Z = \begin{bmatrix} Z_{1,1} & Z_{1,2} & \dots & Z_{1,M} \\ Z_{2,1} & Z_{2,2} & \dots & Z_{2,M} \\ \vdots & \vdots & \ddots & \vdots \\ Z_{N,1} & Z_{N,2} & \dots & Z_{N,M} \end{bmatrix} \quad (۱۱)$$

مدل رگرسیون بردار پشتیبان (SVR)

الگوریتم ماشین بردار پشتیبان (SVM) یکی از روش‌های قدرتمند یادگیری ماشین است که در ابتدا برای طبقه‌بندی توسعه یافت، اما بعدها با گسترش به فرم رگرسیونی (Support Vector Regression) توانست در مدل‌سازی روابط غیرخطی و پیوسته نیز عملکرد بسیار مطلوبی داشته باشد. در حالت رگرسیونی، هدف اصلی SVR یافتن تابعی است که بتواند با کمترین خطا از میان داده‌ها عبور کند، به‌گونه‌ای که اختلاف بین مقادیر مشاهده‌شده و پیش‌بینی‌شده از یک مقدار مجاز (ε) تجاوز نکند. فرم عمومی مدل SVR به‌صورت زیر تعریف می‌شود:

$$f(x) = \langle w, \phi(x) \rangle + b \quad (۱۲)$$

که در آن، w بردار وزن‌ها، b عرض از مبدأ، و $\phi(x)$ تابع نگاشت (Mapping Function) است که داده‌ها را از فضای ورودی به یک فضای ویژگی با بُعد بالاتر انتقال می‌دهد. هدف SVRکمینه‌سازی تابع زیر است:

$$\min_{w,b,\xi_i,\xi_i^*} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \quad (۱۳)$$

با قیود:

$$\begin{cases} y_i - \langle w, \phi(x_i) \rangle - b \le \varepsilon + \xi_i^* \\ \langle w, \phi(x_i) \rangle + b - y_i \le \varepsilon + \xi_i \\ \xi_i, \xi_i^* \ge 0 \end{cases} \quad (۱۴)$$

در این روابط، ε حد مجاز خطا و C پارامتر تنظیمی است که میزان مجازات برای خطاهای بزرگ‌تر از ϵ را تعیین می‌کند. انتخاب تابع کرنل نقش مهمی در عملکرد SVR دارد. در این پژوهش از کرنل شعاعی پایه (RBF) استفاده شده است که قابلیت مدل‌سازی روابط غیرخطی و پیچیده بین متغیرها را فراهم می‌کند و برای پیش‌بینی ضریب دبی (C_d) و بر اساس متغیرهای ورودی استفاده شد. پارامترهای کلیدی شامل C، γ و ε از طریق اعتبارسنجی و بهینه‌سازی تجربی تنظیم شدند تا بهترین تطابق بین داده‌های مشاهده‌شده و مقادیر پیش‌بینی‌شده حاصل گردد. نتایج نشان داد که SVR توانایی بالایی در مدل‌سازی روابط غیرخطی و رفتارهای پیچیده جریان دارد.

مدل Extreme Gradient Boosting (XGBoost)

مدل XGBoost یکی از پیشرفته‌ترین و پرکاربردترین الگوریتم‌های یادگیری تجمیعی (Ensemble Learning) است که بر پایه‌ی چارچوب افزایش گرادیان (Gradient Boosting) توسعه یافته است (Szczepanek, 2022). این الگوریتم با ترکیب تدریجی مجموعه‌ای از درخت‌های تصمیم (Decision Trees)به‌صورت افزایشی، مدلی قدرتمند می‌سازد که در هر مرحله سعی می‌کند خطاهای پیش‌بینی مدل‌های قبلی را اصلاح کند (Piraei et.al, 2023). در واقع، XGBoost با به‌کارگیری فرآیند گرادیان نزولی در فضای تابع (Functional Gradient Descent)، بهینه‌ترین تابع پیش‌بینی را در بین مجموعه‌ای از توابع پایه پیدا می‌کند. در حالت کلی، پیش‌بینی مدل در مرحله‌ی tاُم به‌صورت زیر تعریف می‌شود:

$$\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + \eta f_t(x_i) \quad (۱۵)$$

که در آن، $\hat{y}_i^{(t)}$ مقدار پیش‌بینی‌شده برای داده‌ی i-ام در مرحله‌ی t، $f_t(x)$ درخت تصمیم ساخته‌شده در همان مرحله، و η نرخ یادگیری (Learning Rate) است که میزان تأثیر هر درخت را بر مدل نهایی کنترل می‌کند. الگوریتم در هر تکرار با کمینه‌سازی تابع هدف زیر آموزش داده می‌شود:

$$\Gamma^{(t)} = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right) + \Omega(f_t) \quad (۱۶)$$

در این رابطه، $l(y, \hat{y})$ تابع زیان است که اختلاف بین مقادیر مشاهده‌شده و پیش‌بینی‌شده را اندازه‌گیری می‌کند. عبارت $\Omega(f_t)$ نیز جمله‌ی تنظیم‌کننده است که پیچیدگی مدل را کنترل کرده و از بیش‌برازش جلوگیری می‌کند. این جمله به‌صورت زیر تعریف می‌شود.

$$\Omega(f_t) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2 \quad (۱۷)$$

مدل Category Boosting (CatBoost)

مدل CatBoost یکی از الگوریتم‌های پیشرفته‌ی یادگیری ماشین از خانواده‌ی Gradient Boosting است که به‌طور خاص برای کار با داده‌های پیچیده و غیرخطی توسعه یافته است.این مدل مانند سایر مدل‌های افزایش گرادیان، با ترکیب تدریجی مجموعه‌ای از درخت‌های تصمیم ضعیف مدلی قوی می‌سازد، اما ویژگی بارز آن، استفاده از تکنیک‌های پیشرفته برای جلوگیری از بیش‌برازش و بهبود پایداری آموزش است(Elshaarawy, 2025). این مدل در حالت رگرسیونی، به‌صورت افزایشی ساخته می‌شود تا تابع زیان به حداقل برسد. در حالت رگرسیون، این مدل تلاش می‌کند تابعی مانند F(x) را پیدا کند که تفاوت بین مقادیر واقعی (y) و مقادیر پیش‌بینی شده $(\hat{y})$ را به حداقل برساند.

$$F(x) = \sum_{i=1}^{T} \eta . h_t(x_i) \quad (۱۸)$$

در این معادله X بردار ورودی، T: تعداد کل درخت‌ها ، η: نرخ یادگیری است که تعیین می‌کند هر درخت چقدر در نتیجه نهایی سهم داشته باشد و $h_t(x)$ خروجی درخت شماره t است. این فرآیند از یک ساختار ریاضی پیروی می‌کند که در ادامه به اجزای آن پرداخته می‌شود.

۱-تابع هدف: هدف اصلی کمینه کردن تابع زیان زیر است:

$$Loss(y, \hat{y}) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_j - \hat{y}_i)^2} \quad (۱۹)$$

علاوه بر این، برای کنترل پیچیدگی مدل و جلوگیری از بیش‌برازش یک تابع جریمه نیز به تابع هدف اضافه می‌شود:

$$Obj = \sum L(y_i, \hat{y}_i) + \Omega(f) \quad (۲۰)$$

که در آن $\Omega(f)$ جریمه مربوط به تعداد برگ‌های درخت و مقادیر آن‌هاست.

۲- فرآیند افزایشی: پیش‌بینی نهایی را به‌صورت مجموعه‌ای از درخت‌های کوچک می‌سازد. برای گام t ام به‌صورت زیر است:

$$\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + \eta . h_t(x_i) \quad (۲۱)$$

$\hat{y}_i^{(t-1)}$ پیش‌بینی تجمعی تا مرحله قبلی، η نرخ یادگیری که اثر هر درخت جدید را کنترل می‌کند، $h_t(x_i)$ درخت جدیدی که در مرحله t آموزش دیده تا خطای مرحله قبل را جبران کند.

۳- محاسبه گرادیان: برای اینکه مدل بداند درخت جدید (h_t) را چگونه بسازد، از مشتق تابع زیان نسبت به پیش‌بینی قبلی استفاده می‌کند. این مقدار را باقیمانده می‌نامیم:

$$h(x) = \sum_{j=1}^{2^d} w_j . 1(x \in Leaf_j) \quad (۲۲)$$

d: عمق درخت ، w_j: مقدار خروجی در برگ j ام. برای هر نمونه i، مدل فقط از نمونه‌هایی استفاده می‌کند که در یک ترتیب تصادفی قبل از i قرار گرفته‌اند:

$$\Delta y_i \approx avg(y_i - \hat{y}_j) for \sigma(j) < \sigma(i) \quad (۲۳)$$

این یعنی برای تخمین ضریب دبی سرریز، مدل در هر مرحله سعی می‌کند با استفاده از دانش به دست آمده از شبیه‌سازی‌های قبلی، خطای شبیه‌سازی فعلی را اصلاح کند بدون اینکه داده‌های تکراری باعث فریب مدل شوند.

جدول ۲. شرح متغیرهای مورد استفاده در تحقیق پیشنهادی برای مدل بیان ژن

Table2. Description of the variables used in the proposed investigation to formulate GEP

Variables	Description	Values
G1	Functional set	+, -, *, /, Inv, Sqrt, 3Rt
G2	Chromosomes Number	30
G3	Head Size	8
G4	Number of Genes	3
G5	Size of Gene	32
G6	Connecting Function	+
G7	Fitness Function	RMSE
G8	Inversion Rate	0.00546
G9	Mutation Rate	0.00138
G10	One-point recombination Rate	0.00277
G11	Two-point recombination Rate	0.00277
G12	Stumbling Mutation	0.00141
G13	Gene recombination Rate	0.00277
G14	Root insertion sequence transposition Rate	0.00546
G15	Gene Transportation Rate	0.00277
G16	Uniform Gene Recombination	0.00755

الگوریتم بیان ژن (GEP)

مدل بیان ژن نوعی الگوریتم تکاملی است که ترکیبی از قابلیت‌های الگوریتم ژنتیک (GA) و برنامه‌نویسی ژنتیکی(GP) محسوب می‌شود. مزیت اصلی این مدل نسبت به سایر روش‌های یادگیری ماشین، تولید مدل‌های صریح ریاضی و قابل تفسیر است که توانایی شناسایی روابط غیرخطی پیچیده میان متغیرها را فراهم می‌سازد. با وجود مزایایی همچون دقت و انعطاف‌پذیری بالا، این روش نیازمند تنظیم دقیق پارامترها بوده و در برخی موارد مستعد تولید مدل‌های پیچیده و بیش‌برازش است. مقادیر پارامترهای مورد استفاده از جمله اندازه جمعیت، تعداد نسل و توابع مورد استفاده در این مطالعه در جدول (۲) آمده است. با توجه به دقت روش الگوریتم ژنتیک از نرم‌افزار GeneXproTools به‌منظور تخمین روابط تجربی، از نرم‌افزار مذکور برای تعیین رابطه بین پارامترهای استفاده شد.

پیش‌پردازش و تحلیل داده‌ها

پیش از آموزش مدل‌ها، جهت اطمینان از صحت نتایج آزمایشگاهی، شناسایی داده‌های پرت با استفاده از روش نمره استاندارد (Z-score) انجام شد. در این راستا، نمونه‌هایی که مقدار Z-score آن‌ها خارج از بازه (۳- ، ۳) بود، به عنوان داده‌های ناهنجار شناسایی و حذف گردیدند. در مرحله بعد، اگرچه مدل‌های مبتنی بر درخت شامل XGBoost، CatBoost و RF به طور ذاتی نسبت به مقیاس متغیرها حساس نیستند، اما به منظور بهبود عملکرد مدل ترکیبی Stacked Regression و تسهیل همگرایی در فرآیند بهینه‌سازی توابع هدف، تمامی داده‌های ورودی استانداردسازی شدند. این فرآیند با استفاده از رابطه (۲۴) نجام شد. این اقدام علاوه بر ایجاد بستر مقایسه‌ای یکسان برای تمامی ویژگی‌ها، از سوگیری احتمالی متا مدل در ساختار Stacking جلوگیری کرده و دقت تخمین ضریب دبی (C_d) را افزایش می‌دهد.

$$z = \frac{x-\mu}{\sigma} \quad (۲۴)$$

در این رابطه x مقدار داده مدنظر، μ میانگین کل داده‌ها و σ انحراف معیار داده‌ها است.

شاخص‌های ارزیابی

به‌منظور ارزیابی کمی عملکرد مدل‌های رگرسیونی توسعه داده‌شده، از چهار شاخص اصلی شامل ریشه میانگین مربعات خطا (RMSE)، میانگین خطای مطلق (MAE)، ضریب تبیین (R^2) و شاخص پراکندگی(SI) و کارایی نش–ساتکلیف (NSE) استفاده شد. این شاخص‌ها ترکیب مناسبی از معیارهای دقت عددی، ارزیابی نسبی خطا، توان تبیینی مدل و پایداری عملکرد را فراهم می‌کنند. هرچه مقادیر RMSE، MAE وSI پایین‌تر، NSE و R^2 بالاتر نشان‌دهنده عملکرد بهتر مدل در پیش‌بینی مقادیر تجربی هستند.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_{mea}^{i} - y_{pre}^{i}\right)^2} \quad (۲۵)$$

$$MAE - \frac{100}{N}\sum_{i=1}^{n}\left|y_{mea}^{i} - y_{pre}^{i}\right| \quad (۲۶)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_{mea}^{i} - y_{pre}^{i}\right)^2}{\sum_{i=1}^{n}\left(y_{mea}^{avg} - y_{pre}^{i}\right)^2} \quad (۲۷)$$

$$NSE = 1 - \frac{\sum_{i=1}^{n}\left(y_{mea}^{avg} - y_{pre}^{i}\right)^2}{\sum_{i=1}^{n}\left(y_{mea}^{i} - y_{pre}^{avg}\right)^2} \quad (۲۸)$$

$$SI = \frac{RMSE}{\frac{1}{n}\sum_{i=1}^{n} y_i} \quad (۲۹)$$

که در آن y_i مقادیر مشاهداتی، y^i_{pre} و y^i_{mea} به ترتیب مقادیر تخمینی و اندازه‌گیری شده هستند. y^{avg}_{mea} و n به ترتیب مقدار میانگین مربوط به پارامترهای اندازه‌گیری شده و تعداد اندازه‌گیری‌ها را نشان می‌دهد.

نتایج و بحث

همان‌طور که اشاره شد، هدف از این مطالعه تخمین ضریب سرریز (C_d) در سرریزهای لبه پهن گابیونی با استفاده از مدل‌های جنگل تصادفی، رگرسیون انباشته، ماشین بردار پشتیبان، XGBoost و CatBoost مبتنی بر رگرسیون است. در مطالعه حاضر، در مجموع ۲۳۰ آزمایش اجرا شد که مجموع داده‌های مورد استفاده در بخش مدل‌سازی ۹۲۰ مورد است. با توجه به حساسیت مدل ها به داده‌های ورودی و ویژگی‌ها، اهمیت نسبی ویژگی‌های ورودی برای مدل‌های مختلف برای تخمین ضریب دبی در شکل (۴) ارائه شده است. همان‌طور که مشخص است، دانه‌بندی (d_{50}) و تخلخل ذرات (φ) حساسیت کمتری در تابع هدف مدنظر دارند لذا از متغییرهای ورودی حذف شدند. سپس ضریب سرریز (C_d) در هر یک از مدل‌ها استخراج شد. با استفاده از نتایج به‌دست‌آمده، ضریب سرریز با مدل‌های هوشمند تخمین زده شد. سپس، عملکرد هر یک از مدل‌های هوشمند فوق در تخمین ضریب سرریز با استفاده از پارامترهای مؤثر مورد ارزیابی قرار گرفت.

در این پژوهش، پارامتر C_d به عنوان متغیر هدف (خروجی) مدل انتخاب شد. به منظور ارزیابی حساسیت مدل‌ها نسبت به حجم داده‌های ورودی و انتخاب بهینه‌ترین ساختار برای آموزش، مجموعه داده‌های آزمایشگاهی در چهار سناریوی مختلف تقسیم‌بندی گردید. در این سناریوها، نسبت‌های متفاوتی از کل داده‌ها (شامل ۵۰، ۶۰، ۷۰ و ۸۰ درصد) برای مرحله آموزش در نظر گرفته شد و مابقی داده‌ها جهت اعتبارسنجی و آزمایش مدل اختصاص یافت. فرآیند آموزش و آزمون در هر سناریو به صورت مجزا اجرا شد تا تأثیر حجم داده بر دقت پیش‌بینی مدل‌ها سنجیده شود. انتخاب داده‌ها برای آموزش به صورت تصادفی انتخاب می‌شد. نتایج حاصل از این ارزیابی‌ها بر اساس شاخص‌های آماری در جدول (۳) ارائه شده است.

مقادیر بهینه هایپرپارامترها در مدل‌ها براساس روش grid search به‌دست آمد. به‌طوری که تعداد درختان در شبکه RF برابر ۳۰۰ و عمق چهار، مدل SVR با ε برابر ۰/۱ با کرنل rbf، مدل CatBoost با نرخ یادگیری و عمق درختان و ساختار انتخاب تصادفی در شبکه به ترتیب برابر با ۰/۰۸، ۶ و ۰/۶۵، و مدل XGBoost با نرخ یادگیری ۰/۳۵ عمق ۱۹۴ می‌باشد.

باتوجه به نتایج خروجی مدل‌ها در جدول (۳) دو مدل XGBoost و CatBoost در حالت تخصیص ۸۰ درصد داده‌ها به آموزش نسبت به سایر مدل‌ها با شاخص R^2، به ترتیب ۰/۹۶۰ و ۰/۹۶۲ بهترین عملکرد را دارند. بنابراین دو مدل XGBoost و CatBoost توانسته‌اند نوسانات و رفتار داده‌های آزمایشگاهی مربوط به ضریب دبی را با دقت بالاتری پیش‌بینی کنند، که علت نوسانات انتخاب تصادفی داده است. تفاوت مدل با داده‌های آزمایشگاهی بسیار کم و داده‌ها بسیار منطبق بر یکدیگر هستند. فقط در چند نقطه محدود، اختلافات جزئی مشاهده می‌شود که قابل‌چشم‌پوشی است. شکل (۵) مقایسه شاخص NSE در نسبت‌های مختلف تقسیم‌بندی داده‌های شبکه را نشان می‌دهد، که مدل‌های XGBoost و CatBoost دقت مناسب‌تری نسبت به سایر مدل‌ها دارند.

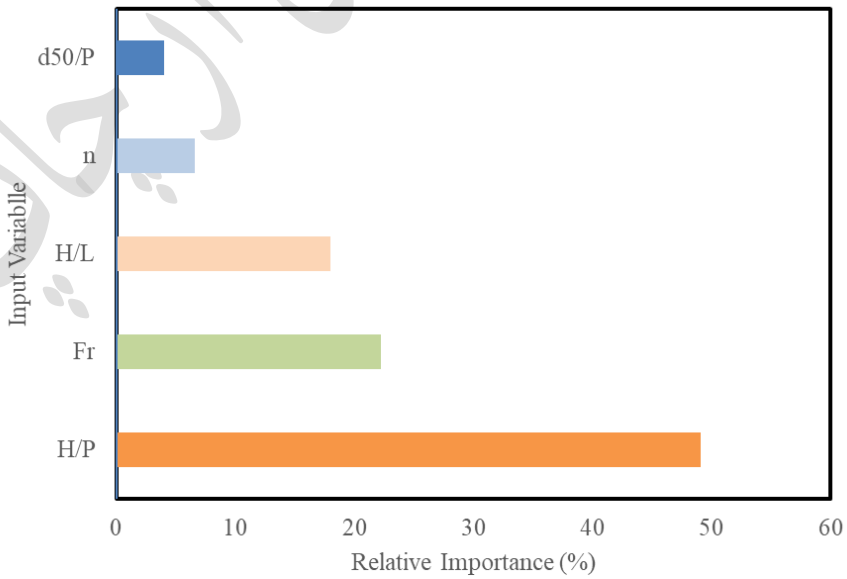


Fig. 4- Sensitivity analysis of input features with respect to the objective function

شکل ٤- تحلیل حساسیت ویژگی‌های ورودی نسبت به تابع هدف

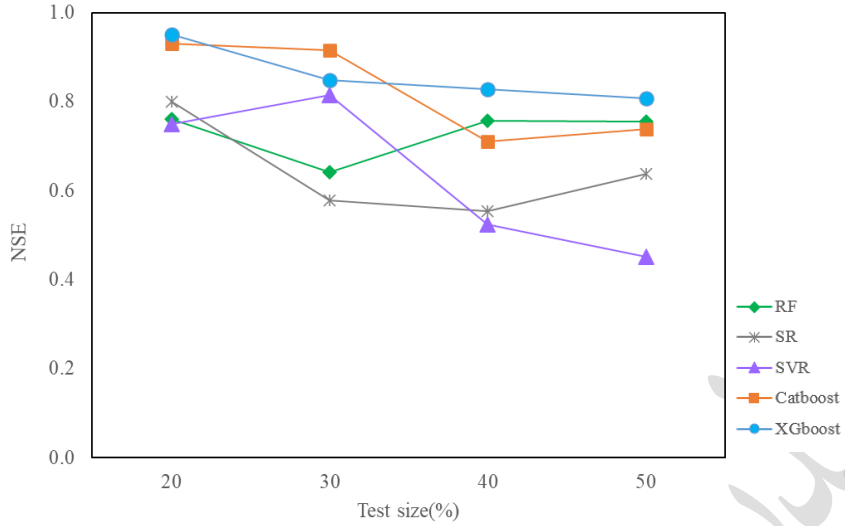


Fig. 5 - Comparison of Nash-Sutcliffe index in different network data partitioning ratios

شکل ۵- مقایسه شاخص نش-ساتکلیف در نسبت‌های مختلف تقسیم‌بندی داده‌های شبکه

جدول ۳- شاخص‌های ارزیابی مدل‌های مختلف مورد استفاده

Table 3- Evaluation indicators of different models used

·|·	Network size %	All				Train				Test			
		R^2	MAE	RMSE	SI	R^2	MAE	RMSE	SI	R^2	MAE	RMSE	SI
RF	50-50	0.807	0.057	0.144	0.160	0.934	0.029	0.082	0.093	0.690	0.083	0.186	0.202
	60-40	0.903	0.045	0.102	0.113	0.951	0.032	0.076	0.081	0.791	0.063	0.131	0.156
	70-30	0.915	0.040	0.095	0.106	0.939	0.032	0.084	0.093	0.843	0.058	0.117	0.132
	80-20	0.937	0.034	0.082	0.091	0.945	0.029	0.078	0.087	0.898	0.083	0.095	0.104
SR	50-50	0.771	0.173	0.156	0.173	0.821	0.072	0.142	0.277	0.712	0.108	0.168	0.192
	60-40	0.856	0.137	0.124	0.137	0.892	0.054	0.107	0.225	0.801	0.080	0.145	0.159
	70-30	0.890	0.120	0.108	0.120	0.889	0.058	0.119	0.133	0.893	0.052	0.076	0.088
	80-20	0.895	0.161	0.145	0.161	0.890	0.092	0.152	0.213	0.899	0.084	0.116	0.136
SVR	50-50	0.646	0.092	0.209	0.232	0.918	0.061	0.084	0.096	0.563	0.124	0.284	0.310
	60-40	0.831	0.090	0.155	0.173	0.872	0.081	0.156	0.171	0.743	0.102	0.154	0.175
	70-30	0.846	0.078	0.140	0.155	0.875	0.076	0.130	0.144	0.787	0.085	0.160	0.180
	80-20	0.851	0.076	0.147	0.164	0.834	0.076	0.154	0.168	0.882	0.080	0.116	0.139
XGBR	50-50	0.928	0.033	0.089	0.103	0.980	0.002	0.005	0.005	0.812	0.065	0.126	0.144
	60-40	0.925	0.048	0.092	0.102	0.984	0.031	0.047	0.052	0.833	0.073	0.134	0.148
	70-30	0.952	0.026	0.071	0.079	0.983	0.014	0.046	0.051	0.827	0.054	0.110	0.125
	80-20	0.990	0.019	0.038	0.042	0.994	0.013	0.026	0.029	0.960	0.041	0.065	0.076
CBR	50-50	0.828	0.031	0.135	0.150	0.997	0.001	0.0005	0.020	0.716	0.062	0.190	0.212
	60-40	0.976	0.021	0.050	0.056	0.998	0.009	0.0140	0.015	0.893	0.038	0.077	0.087
	70-30	0.913	0.031	0.096	0.106	0.994	0.014	0.0228	0.026	0.799	0.070	0.171	0.178
	80-20	0.996	0.008	0.019	0.022	0.999	0.004	0.0054	0.006	0.962	0.026	0.042	0.05

باتوجه به شکل (۶) و جدول (۳) با درنظر گرفتن نسبت آموزش روی کل داده‌ها، مدل‌های XGBoost و CatBoost تمام نقاط داده در سطح رضایت‌بخشی از دقت ±۱۰٪ قرار می‌گیرند. مدل XGBoost از نظر تمام معیارهای دقت، با اختلاف ۲/۰ درصد در مقدار R^2، از CatBoost بهتر عمل می‌کند. در نتیجه، می‌توان نتیجه گرفت که اگرچه مدل CatBoost از سایر مدل‌ها بهتر عمل کرده است، اما XGBoost با افزایش ۲/۰ درصد مقدار R^2 و کاهش تقریباً ۲۵ درصدی مقدار RMSE، عملکرد قابل توجهی نسبت به XGBoost داشته است. همچنین برای آموزش ۸۰ درصد داده‌ها مدل‌های CatBoost و XGBoost با شاخص آمار NSE به‌ترتیب برابر ۰/۹۳ و ۰/۹۵ و برای مدل‌های RF، SR، SVR به‌ترتیب ۰/۷۶، ۰/۸۰، ۰/۷۴ می‌باشد.

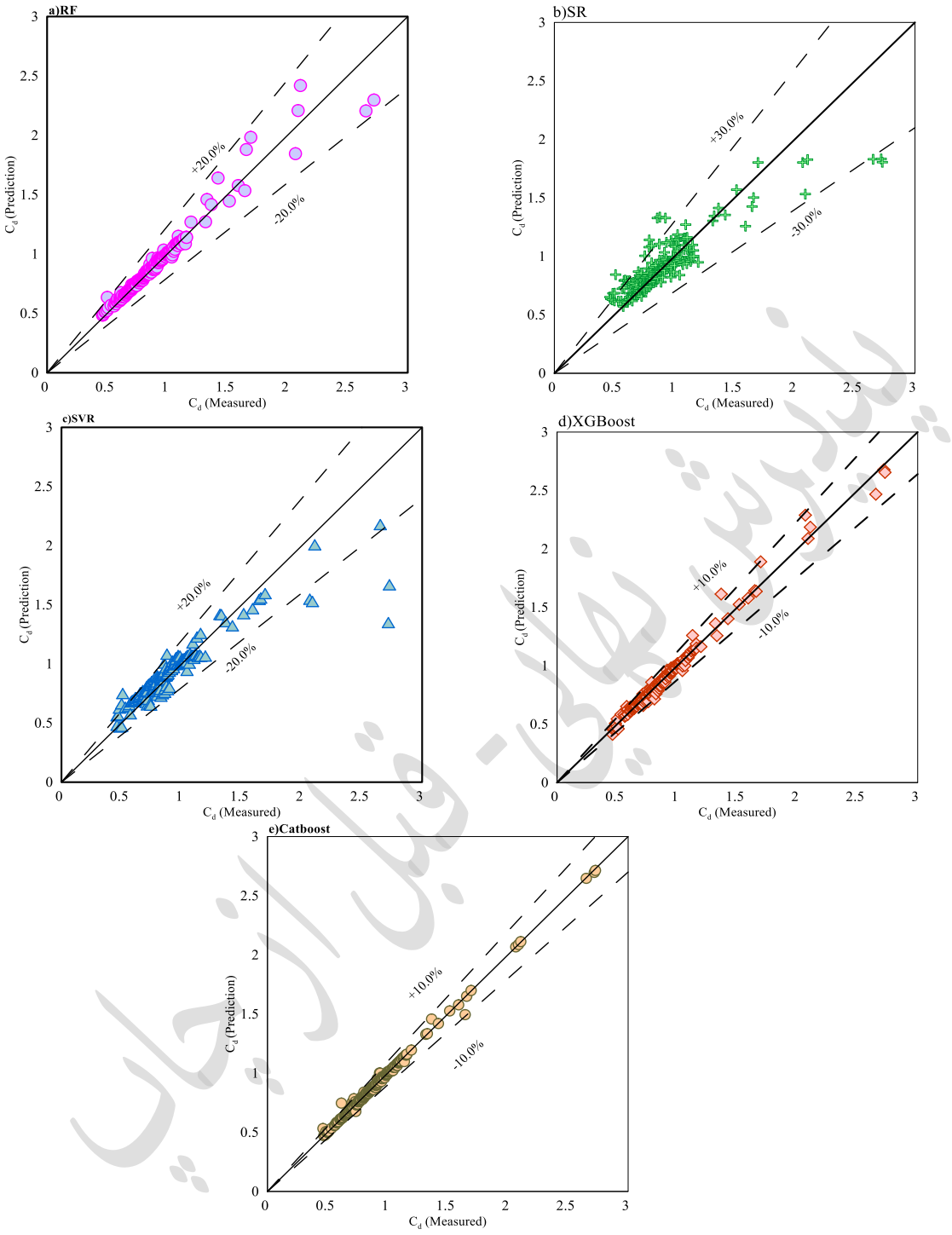


Fig. 6- Results of training of the considered models with the (80-20) split in C_d prediction

شکل ٦- نتایج آموزش مدل‌های مدنظر برای کل داده ها در پیش‌بینی C_d

به‌منظور ارزیابی دقیق و مقایسه کارایی مدل‌های توسعه‌یافته RF، SR، SVR، XGBoost و CatBoost در پیش‌بینی ضریب دبی (C_d)، نمودارهای پراکندگی مقادیر اندازه‌گیری‌شده در برابر مقادیر پیش‌بینی‌شده برای فاز آموزش (۸۰ درصد داده‌ها) و تست (۲۰ درصد داده‌ها) در شکل (۷) ارائه شده است. بر اساس میزان پراکندگی نقاط توانایی تعمیم‌پذیری و عملکرد مدل‌ها در مواجهه با داده‌های حدی، می‌توان کارایی مدل‌های مورد مطالعه را دسته‌بندی کرد. مدل CatBoost پایدارترین و دقیق‌ترین رفتار محاسباتی را در کل پهنای داده‌ها به نمایش گذاشته است. انطباق کامل و فشرده نقاط مربوط به هر دو فاز آموزش و تست، نشان‌دهنده ضریب تبیین R^2 بسیار بالا و کمترین میزان خطای محاسباتی است. نکته حائز اهمیت در عملکرد این مدل، توانایی ممتاز آن در پیش‌بینی دقیق مقادیر حدی بدون بروز پدیده بیش‌برازش یا کم‌برآوردی در مرزها است. مدل XGBoost جهش چشمگیری را در دقت پیش‌بینی نشان می‌دهد. در این الگوریتم، انحراف داده‌های آموزش تقریباً به صفر رسیده و کاملاً بر خط فرضی مماس شده‌اند. همچنین، داده‌های فاز آزمون با تراکم بسیار بالا و خطای ناچیز حول خط تعادل توزیع شده‌اند که قدرت تعمیم‌پذیری بالای مدل را تأیید می‌کند. مدل RF در شرایط آزمون دچار انحراف جزئی و کم‌برآوردی می‌شود که این امر ناشی از گرایش ذاتی درخت‌های تصادفی مجزا به سمت میانگین داده‌ها در نقاط مرزی است؛ اگرچه ساختار کلی آن در مقایسه با SVR پایدارتر است.

مدل SVR در بازه داده‌های کوچک تا متوسط عملکرد قابل قبولی دارد، اما با افزایش پارامتر هدف ($C_d>1$) عبور از محدوده، نتایج فاز تست دچار انحراف خطی موازین شده و تا حدودی روند صعودی واقعی داده‌ها را گم می‌کنند که نشان‌دهنده حساسیت بالای این الگوریتم به تنظیمات پارامترهای هسته در مواجهه با روابط شدیداً غیرخطی است. در نهایت، مدل رگرسیون انباشته (SR) ضعیف‌ترین کارایی را در میان تمامی روش‌های بررسی‌شده به ثبت رسانده است. الگوریتم انباشته در فاز تست در مواجهه با مقادیر C_d کاملاً دچار پهن‌شدگی افقی و پراکندگی شدید شده است. این رفتار ثابت می‌کند که مدل ترکیب رگرسیونی در نگاشت و تعمیم‌دهی پیچیدگی‌های هیدرولیکی حاکم بر ضریب دبی در جریان‌های شدید و شرایط حدی ناتوان است .همچنین می‌توان نتیجه گرفت که پیش‌بینی داده‌ها در هر مدل به نسبت تقسیم‌بندی داده‌ها برای آموزش و آزمون وابستگی دارند. عملکرد برتر XGBoost را می‌توان به توانایی این مدل در استخراج برهم‌کنش‌های غیرخطی میان پارامترهای هندسی سرریز و ویژگی‌های جریان نسبت داد؛ موضوعی که در سرریزهای متخلخل اهمیت ویژه‌ای دارد.

شکل (۸) نمودار ویولنی مقادیر پیش‌بینی‌شده ضریب دبی (C_d) را توسط پنج الگوریتم یادگیری ماشین به تصویر کشیده است. این نمودار به طور همزمان چگالی توزیع فراوانی و محدوده گسترش داده‌های پیش‌بینی‌شده را نشان می‌دهد. همان‌طور که مشاهده می‌شود، بخش عمده‌ای از بدنه چگالی تمامی مدل‌ها (قسمت عریض ویولن‌ها) در محدوده ۰/۵ تا ۱/۲ متمرکز شده است که نشان‌دهنده همگرایی بالای الگوریتم‌ها در بازه متراکم داده‌های آزمایشگاهی است. با این حال، تفاوت اصلی مدل‌ها در پهنای دم بالایی آن‌ها نمایان می‌شود؛ جایی که دو مدل CatBoost و XGBoost با کشیدگی دم تا مقادیر بالاتر از ۲/۵، پتانسیل ممتاز خود را در ردیابی و بازسازی توزیع مقادیر حدی و سقف ضریب دبی اثبات می‌کنند. مدل‌های SVR و SR با داشتن دم‌های کوتاه‌تر، محدودیت ساختاری خود را در بازتولید مقادیر بزرگ نشان می‌دهند و تمایل دارند مقادیر پیش‌بینی‌شده را به سمت میانگین داده‌ها فشرده کنند. این تحلیل توزیعی، یافته‌های حاصل از نمودارهای پراکندگی مبنا بر برتری مطلق مدل‌های تقویت گرادیان را کاملاً تایید می‌کند.دیاگرام تیلور، نتایج به‌دست‌آمده از جدول (۳) شاخص‌های آماری را تأیید و بصری‌سازی می‌کند.

باتوجه به شکل (۹) تحلیل تطبیقی در فضای تیلور همسو با نتایج شاخص‌های MAE و RMSE نشان می‌دهد. مدل XGBoost با ثبت همبستگی بیشتر از ۰/۹۹، انحراف معیار تطبیق‌یافته ۰/۳۲ و کمترین خطا، به عنوان مدل بهینه، مکانیزم‌های پیچیده جریان در سرریز گابیونی را با بالاترین شبیه‌سازی ارایه می‌کند؛ این مدل نه تنها در شاخص‌های MAE و RMSEکمترین مقادیر را ثبت کرده، بلکه در نمودار تیلور نیز با بازتولید دقیق انحراف معیار مشاهداتی و ثبت بالاترین میزان همبستگی، پایداری و قدرت تعمیم‌پذیری بالای خود را در مقایسه با مدل رگرسیون انباشته (SR) به اثبات رسانده است که گویای دقت محاسباتی فوق‌العاده آن در تخمین ضریب دبی سرریز است. در سمت مقابل، مدل‌های CatBoost با چالش بیش‌برآورد واریانس و RF با ثبت همبستگی ضعیف‌تر و دقت متوسط در رتبه‌های بعدی قرار می‌گیرند، در حالی که مدل SR به دلیل ناتوانی شدید در شبیه‌سازی واریانس و خطای بسیار بالا، کاملاً فاقد کارایی لازم است.

در این پژوهش، مدل XGBoost با توانایی بالا در ارزیابی روابط غیرخطی توانست به ضریب تبیین ۰/۹۹ دست یابد. در مقایسه با مطالعه‌های گذشته، Azma et.al (2023) با استفاده از مدل مذکور به دقت ۰/۹۹۲ رسیده بودند که نشان می‌دهد الگوریتم ساختاریافته‌ی XGBoost در این پژوهش عملکرد بهتری در پیش‌بینی متغیرها داشته است. همچنین، توانایی مدل CatBoost در مقایسه با سایر نتایج پژوهشگران، Azma et.al (2023) و Hassanian Shirvan et.al (2025) که با استفاده از مدل مذکور به‌دقت ۰/۹۹۴ و ۰/۹۹۵ دست یافتند، اما پژوهش حاضر توانست ضریب تبیین ۰/۹۹۶ دست یابد که با توجه به تعداد ویژگی‌های کمتر (ورودی‌های بهینه‌تر) به این کارایی رسیده است که از نظر محاسباتی به‌صرفه‌تر است.

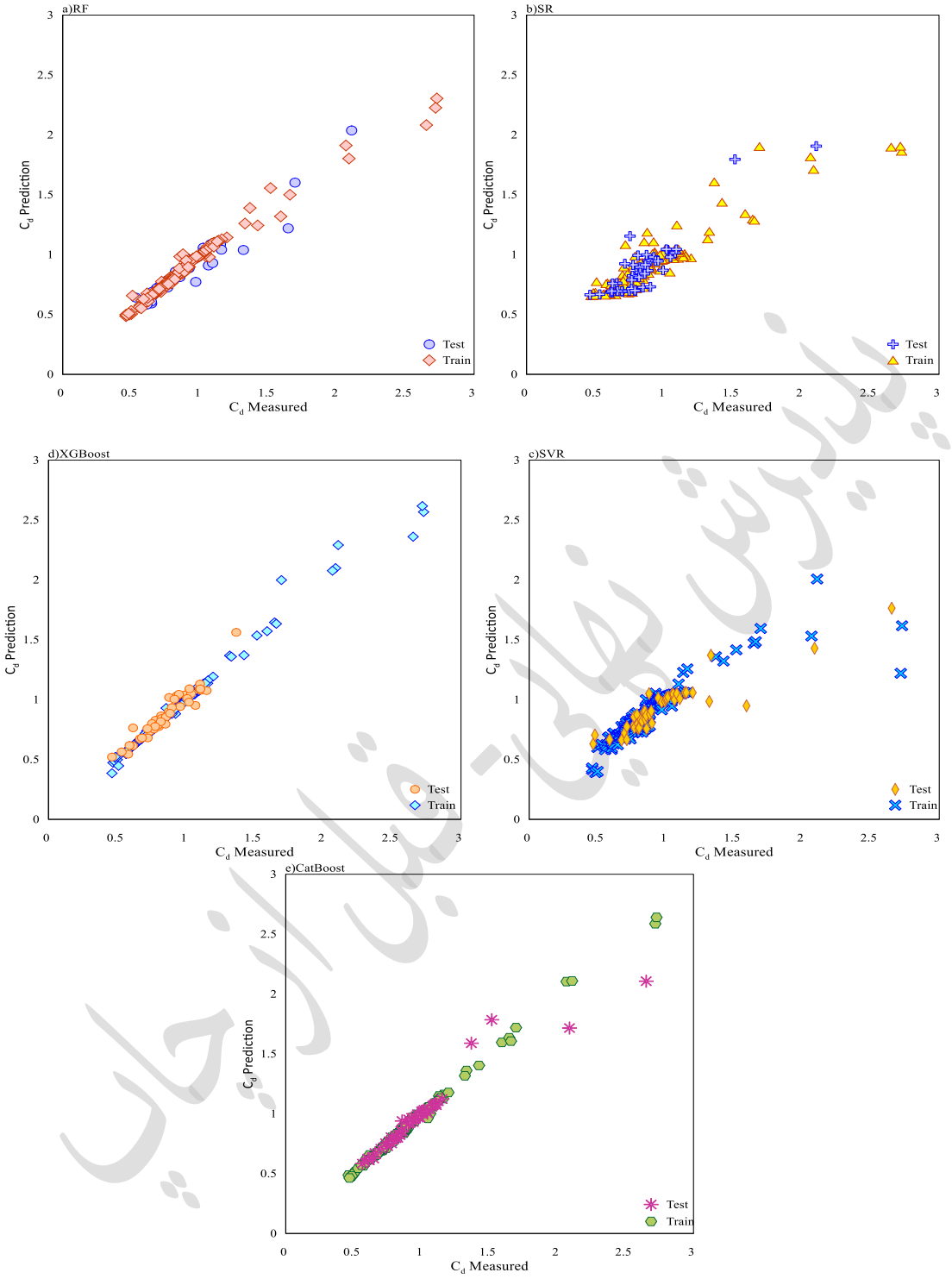


Fig. 7- Results of different classific ations of training the models considered in C_d prediction

شکل ۷- نتایج آموزش مدل‌های مدنظر برای نسبت ۸۰ - ۲۰ در پیش‌بینی C_d

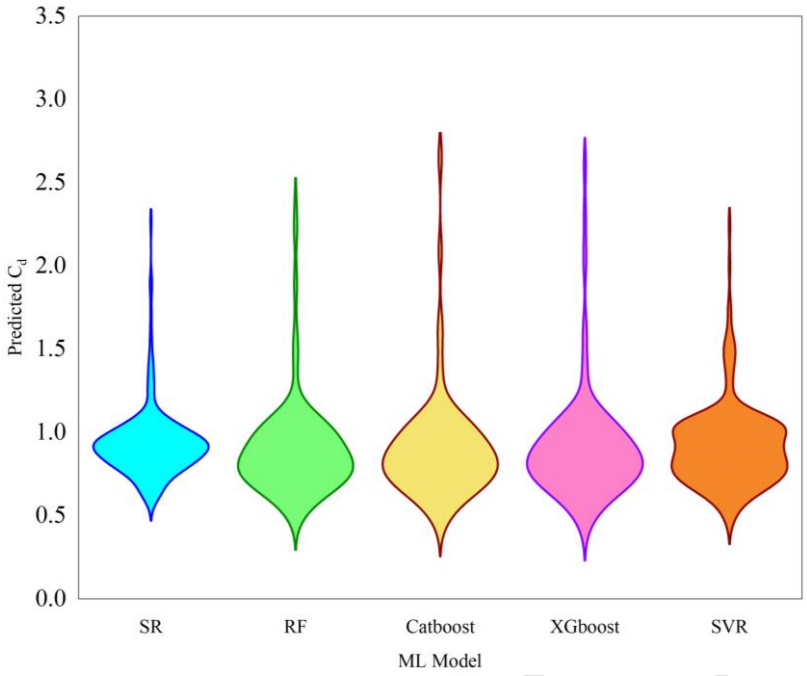


Fig 8. Violin plot of predicted values for the discharge coefficient

شکل ۸ - نمودار ویولنی مقادیر پیش‌بینی‌شده ضریب دبی

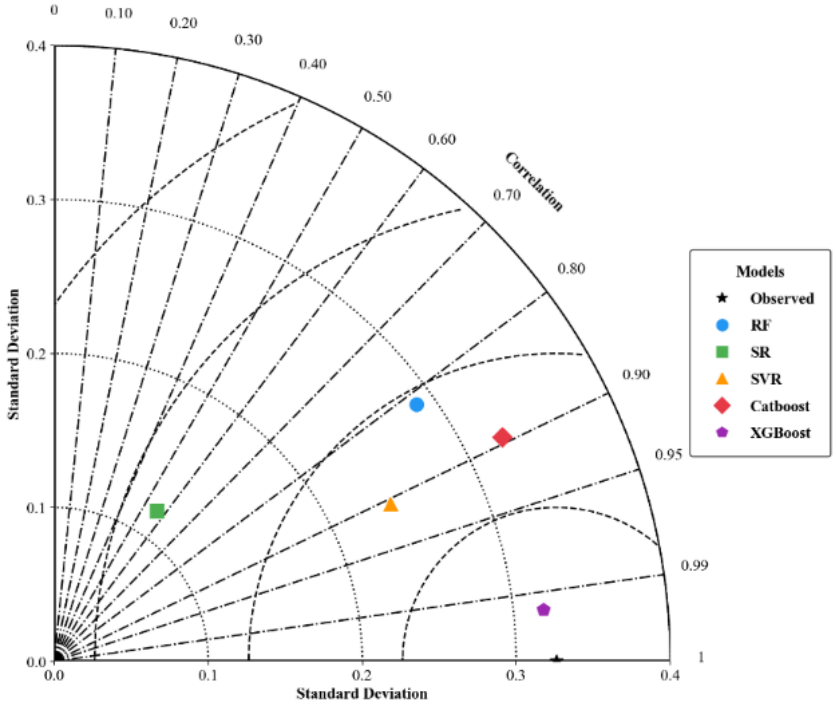


Fig. 9- Taylor diagram illustrating the performance of machine learning models in predicting C_d

شکل۹- دیاگرام تیلور عملکرد مدل‌های پیش‌بینی C_d

در ادامه از نرم‌افزار GeneXproTools با بهره‌گیری از تکنیک برنامه‌نویسی بیان ژن (GEP) به‌کار گرفته شد تا رابطه میان پارامترهای بی‌بعد تعیین گردد. در حالت تخصیص داده‌ها به ۸۰ و ۷۰ درصد آموزش، مقادیر شاخص‌های آماری (R^2)، (RMSE)، MAE برای نسبت ۸۰ درصد به ترتیب ۰/۹۵۵ ، ۰/۱۰۶ و۰/۰۵۵ و برای نسبت ۷۰ درصد به ترتیب ۰/۹۴۲، ۰/۰۸۴ و ۰/۰۶۱

به‌دست آمد. معادله ۲۸ برای نسبت آموزش ۸۰ درصد و معادله (۲۹) برای نسبت آموزش ۷۰ درصد داده‌ها می‌باشد. با تخصیص آموزش به کل داده‌ها در این مدل معادله (۳۰) استخراج شد. تحلیل آماری برای ارزیابی دقت معادلات برای سرریزهای گابیونی لبه‌پهن مستطیلی درشرایط مختلف تخصیص آموزش داده‌ها انجام شد. شکل (۱۰) (a)، عملکرد معادله (۳۲) برای تمامی داده ها آموزش دیده ، شکل (۱۰) (b) عملکرد معادله (۳۱) برای تمامی ۸۰ درصد داده های آموزش دیده و شکل (۱۰) (c)، عملکرد معادله ۳۰ برای ۷۰ درصد داده ها آموزش دیده را نشان می‌دهد که با درجه بالایی از انطباق با خط کامل (خط ۱:۱) قرار گرفته‌اند، به‌طوری‌که تقریباً ۹۵ درصد از مجموعه داده‌ها هنگام محاسبه با استفاده از معادله (۳۲) خطاهایی در محدوده ±۱۵ نشان می‌دهند.

$$C_d = \frac{Fr^3 P}{H} - 1.112\frac{Fr^2 P}{H} + \frac{H^2}{LP} - \frac{H^3}{L^3} + \frac{1}{0.092(\frac{H}{L}) + \frac{H}{PFr}} \quad (۳۰)$$

$$C_d = \frac{0.846\, FrLP}{H(L+P)} + \frac{HFr^2}{L} + \frac{H^2}{LP} - 0.665\frac{H}{L} + \sqrt{\frac{HFr}{L}} + \frac{HLFr}{P(H-2L)} \quad (۳۱)$$

$$C_d = \frac{\frac{H}{L}}{Fr(\frac{P}{H} + 5.898) - \frac{H}{P}} + \frac{H^2(L-P)}{12.109L^2P} - 0.177\frac{H}{L} + Fr + 0.1805\, Fr\left(\frac{P}{H}\right)^{1.5} \quad (۳۲)$$

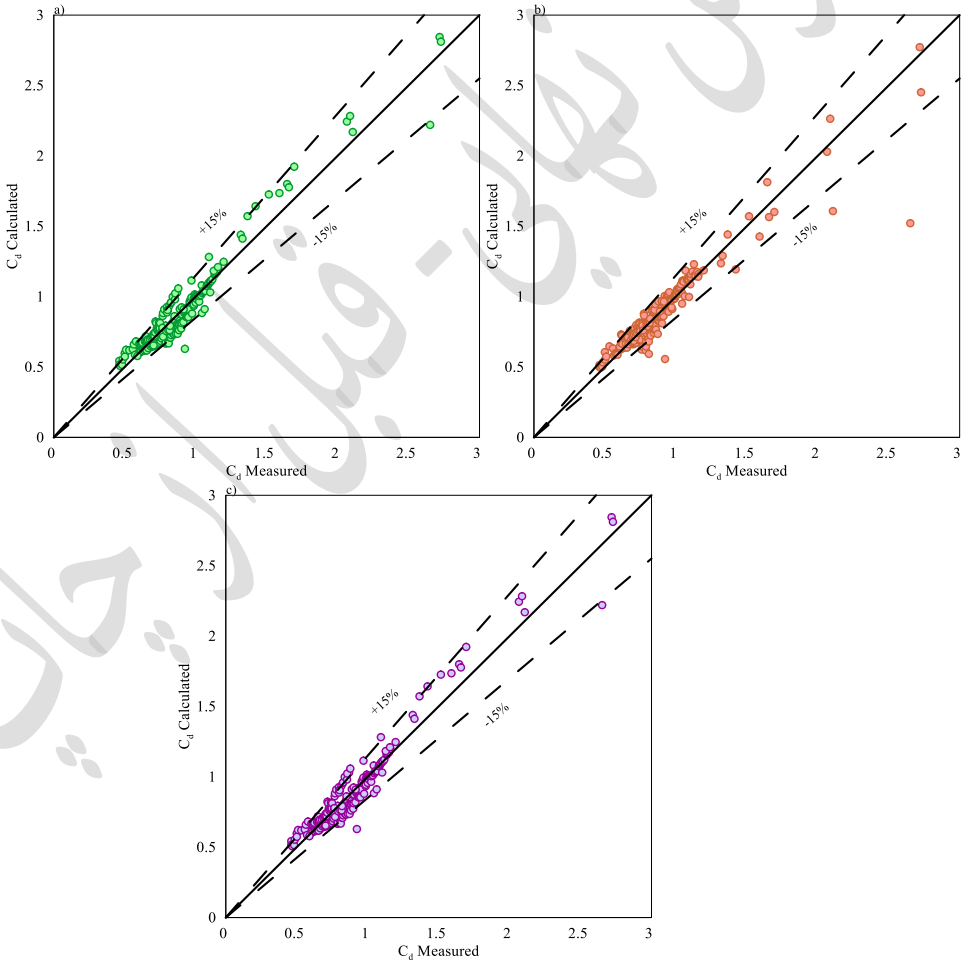


Fig. 10 - Performance of the GEP model for the entire training data in predicting C_d

شکل ۱۰- عملکرد مدل بیان ژن برای کل داده‌های آموزش در پیش‌بینی C_d

نتیجه‌گیری

سرریزهای گابیونی لبه‌پهن مستطیلی به دلیل برخورداری از ویژگی‌هایی نظیر قابلیت اندازه‌گیری دقیق دبی، کنترل مؤثر جریان و کاهش انرژی، مقاومت سازه‌ای بالا و سازگاری زیست‌محیطی، کاربرد گسترده‌ای در کانال‌ها، رودخانه‌ها و سازه‌های هیدرولیکی کوچک تا متوسط دارند. در این پژوهش، کارایی و پایداری پنج مدل هوش مصنوعی مبتنی بر یادگیری ماشین مبتنی بر رگرسیون شامل RF، SR، SVR و مدل‌های تقویت گرادیان پیشرفته XGBoost و CatBoost در قالب چهار سناریوی تخصیص داده ۸۰، ۷۰، ۶۰ و ۵۰ درصد داده های خام جهت آموزش مدل در راستای پیش‌بینی ضریب دبی (C_d) مورد ارزیابی و مقایسه قرار گرفت. یافته‌های کلیدی نشان داد که حجم داده‌های آموزش تأثیر مستقیمی بر توانایی تعمیم‌پذیری مدل‌ها دارد، به‌طوری که سناریوی تفکیک ۸۰–۲۰ بهینه‌ترین بستر را برای استخراج الگوهای غیرخطی حاکم بر جریان هیدرولیکی فراهم نمود. در میان الگوریتم‌های توسعه‌یافته، مدل‌های رگرسیون انباشته و ساختارهای تقویت گرادیان به طور نظام‌مند برتری قاطعی در تخمین ضریب دبی نسبت به رویکردهای سنتی نشان دادند؛ در این میان، مدل‌های XGBoost و CatBoost در سناریوی ۸۰–۲۰ با کسب بالاترین ضریب تبیین به ترتیب ۰/۹۶۰ و ۰/۹۶۲ در فاز آزمون ۰/۹۸۷ و ۰/۹۹۶ در کل داده‌ها و کمترین میزان خطای ساختاری ۰/۰۴۱ و ۰/۰۴۲، پایدارترین عملکرد را بدون مواجهه با پدیده بیش‌برازش ارائه داد. همچنین استخراج معادله میان پارامترهای بی‌بُعد به کمک الگوریتم GEP نیز نشان داد که نسبت آموزش داده‌ها بر تولید یک معادله دقیق تاثیرگذار خواهد بود.

تضاد منافع نویسندگان

نویسندگان این مقاله اعلام می‌دارند که هیچ گونه تضاد منافعی در خصوص نگارش و انتشار مطالب و نتایج این پژوهش ندارند.

منابع مالی

نویسندگان هیچ گونه حمایت مالی برای تحقیق، تالیف و انتشار این مقاله دریافت نکردند

دسترسی به داده‌ها

مجموعه داده‌های تولید شده و یا تحلیل شده در طول مطالعه حاضر، در صورت درخواست معقول، از نویسنده مسئول در دسترس هستند.

مشارکت نویسندگان

نویسنده اول: سید علیرضا اسماعیلی، انجام آزمایش‌ها، تجزیه و تحلیل اولیه داده‌ها، کدنویسی مدل‌های یادگیری ماشین، نگارش مقاله.

نویسنده دوم: مجید حیدری، بررسی نتایج و کدنویسی‌ها، بررسی و اصلاح مقاله، اصلاح نهایی مقاله.

نویسنده سوم: سعید گوهری، بررسی و اصلاح مقاله، اصلاح نهایی مقاله.

تقدیر وتشکر

بدین وسیله نویسندگان مقاله از همکاری‌ها و حمایت‌های مدیر گروه و سرپرست مجموعه آزمایشگاه‌های گروه علوم و مهندسی آب دانشگاه فردوسی مشهد تشکر و قدردانی می‌نمایند.

References

1- Akbari, M., Salmasi, F., Arvanaghi, H., Karbasi, M. and Farsadizadeh, D., 2019. Application of Gaussian process regression model to predict discharge coefficient of gated piano key weir. *Water Resources Management*, *33*(11), pp.3929–3947. doi:10.1007/s11269-019-02343-3.

2- Alomari, N.K., Altalib, A.N. and Al-Janabi, A.M.S., 2023. Discharge estimation using brink depth over a trapezoidal-shaped weir. *Flow Measurement and Instrumentation*, 94, 102454. doi:10.1016/j.flowmeasinst.2023.102454.

3- Azamathulla, H.M., Ghani, A.A., Zakaria, N.A. and Guven, A., 2010. Genetic programming to predict bridge pier scour. *Journal of Hydraulic Engineering*, *136*(3), pp.165–169. doi:10.1061/(ASCE)HY.1943-7900.0000133.

4- Azma, A., Tavakol Sadrabadi, M., Liu, Y., Azma, M., Zhang, D., Cao, Z. and Li, Z., 2023. Boosting ensembles for estimation of discharge coefficient and through flow discharge in broad-crested gabion weirs. *Applied Water Science*, *13*(2), 45. doi:10.1007/s13201-022-01841-x.

5- Biabani, R., Salmasi, F., Nouri, M. and Abraham, J., 2022. Flow over embankment gabion weirs in free flow conditions. *Journal of Hydro-Environment Research*, 44, pp.65–76. doi:10.1016/j.jher.2022.08.001.

6- Doustkam, M., Rahmanshahi, M., Fathi-Moghadam, M., Keramat, A. and Duan, H.F., 2024. Experimental study on the hydraulic performance of porous broad-crested weirs with sloping crests. *Water Resources Management*, *38*(12), pp.4783–4802. doi:10.1007/s11269-024-03887-9.

7- Elshaarawy, M.K., 2025. Metaheuristic-driven CatBoost model for accurate seepage loss prediction in lined canals. *Multiscale and Multidisciplinary Modeling, Experiments and Design*, 8, 235. doi:10.1007/s41939-025-00800-8.

8- Esmaili, S. A., Heydari, M., Gohari, S. Hydrodynamic Analysis of Flow and Discharge Coefficient in Weirs with Porous Media under Free Overtopping Conditions. *Journal of Water and Sustainable Development*, 2026; *12*(4): 88-102. doi: 10.22067/jwsd.v12i4.2508-1449. (In Persian).

9- Haghiabi, A.H., Parsaie, A. and Ememgholizadeh, S., 2018. Prediction of discharge coefficient of triangular labyrinth weirs using Adaptive Neuro Fuzzy Inference System. *Alexandria Engineering Journal*, *57*(3), pp.1773-1782. doi:10.1007/s41939-025-00800-8.

10- Hameed, M.M., Alomar, M.K., Razali, S.F.M. and Salem, A., 2025. Integrated approach of extreme learning machines and locally weighted linear regression for improved discharge coefficient prediction. *Scientific Reports*, *15*(1), 21761. doi:10.1038/s41598-025-03812-z.

11- Hassanin Shirvan, S.H., Pirzadeh, B., Rajaei, S.H. and Shafai Bejestan, M., 2025. Multi-approaches evaluation for prediction of discharge coefficient of porous broad-crested weirs under upstream partial blockage. *Iranian Journal of Science and Technology, Transactions of Civil Engineering*, *49*(5), pp.4913-4925. doi.org/10.1007/s40996-025-01737-1.

12- Heravi, Z., Moghaddam, M.A., Esmaili, K., Givehchi, M. and Bahrpeyma, A., 2024. Impact of installing new scour prevention countermeasure (F-jacks) around bridge pier on flow pattern and scour control. *Ocean Engineering*, *313*, p.119326. https://doi.org/10.1016/j.oceaneng.2024.119326.

13- Jiang, L., Diao, M., Sun, H. and Ren, Y., 2018. Numerical modeling of flow over a rectangular broad-crested weir with a sloped upstream face. *Water*, *10*(11), 1663. doi:10.3390/w10111663.

14- Khatibi, R., Salmasi, F., Ghorbani, M.A. and Asadi, H., 2014. Modelling energy dissipation over stepped-gabion weirs by artificial intelligence. *Water Resources Management*, *28*(7), pp.1807–1821. doi:10.1007/s11269-014-0545-y.

15- Kim, S., Singh, V.P. and Seo, Y., 2014. Evaluation of pan evaporation modeling with two different neural networks and weather station data. *Theoretical and Applied Climatology*, *117*(1), pp.1–13. doi:10.1007/s00704-013-0985-y.

16- Kumar, S., Garg, N., Das, S.K. and Pandey, M., 2025. The Role and Impact of Gabion Weirs in River Management: A Systematic Review. In: Pandey, M., et al. (eds) *Flood Forecasting and Hydraulic Structures*. *Lecture Notes in Civil Engineering*, vol 546. Springer, Singapore. doi:10.1007/978-981-97-9168-2_4.

17- Nourani, B., Arvanaghi, H. and Salmasi, F., 2021. Effects of different configurations of sloping crests and upstream and downstream ramps on the discharge coefficient for broad-crested weirs. *Journal of Hydrology*, 603, 126940. doi:10.1016/j.jhydrol.2021.126940.

18- Nourani, B., Arvanaghi, H., Pourhosseini, F.A., Javidnia, M. and Abraham, J., 2023. Enhanced support vector machine with particle swarm optimization and genetic algorithm for estimating discharge coefficients of circular-crested oblique weirs. *Iranian Journal of Science and Technology, Transactions of Civil Engineering*, *47*(5), pp.3185-3198. doi:10.1007/s40996-023-01110-0.

19- Pallavi, M. and Harashit, L., 2018. Open channel flow characteristics using gabion weir. International Research Journal of Engineering and Technology, *5*(12), pp.928-933.

20- Piraei, R., Afzali, S.H. and Niazkar, M., 2023. Assessment of XGBoost to estimate total sediment loads in rivers. *Water Resources Management*, *37*(13), pp.5289–5306. doi:10.1007/s11269-023-03606-w.

21- Rahmanshahi, M., Jafari-Asl, J., Fathi-Moghadam, M., Ohadi, S. and Mirjalili, S., 2024. Metaheuristic learning algorithms for accurate prediction of hydraulic performance of porous embankment weirs. *Applied Soft Computing*, 151, 111150. doi:10.1016/j.asoc.2023.111150.

22- Rahmanshahi, M. and Shafai Bejestan, M., 2020. Gene-expression programming approach for development of a mathematical model of energy dissipation on block ramps. *Journal of Irrigation and Drainage Engineering*, *146*(2), 04019033. doi:10.1061/(ASCE)IR.1943-4774.0001442.

23- Roushangar, K., Akhgar, S., Salmasi, F. and Shiri, J., 2014. Modeling energy dissipation over stepped spillways using machine learning approaches. *Journal of Hydrology*, 508, pp.254–265. doi:10.1016/j.jhydrol.2013.10.053.

24- Roustaei, N., 2024. Application and interpretation of linear-regression analysis. *Medical Hypothesis, Discovery & Innovation in Ophthalmology*, *13*(3), pp.151–159. doi:10.51329/mehdiophthal1506.

25- Salmasi, F. and Abraham, J., 2020. Discussion of "Hydrodynamics of rectangular broad-crested porous weirs" by Akbar Safarzadeh and Seyed Hossein Mohajeri. *Journal of Irrigation and Drainage Engineering*, *146*(4), 07020003. doi:10.1061/(ASCE)IR.1943-4774.0001450.

26- Salmasi, F., Sattari, M.T. and Nurcheshmeh, M., 2021. Genetic programming approach for estimating energy dissipation of flow over cascade spillways. *Iranian Journal of Science and Technology, Transactions of Civil Engineering*, *45*(1), pp.443–455. doi:10.1007/s40996-020-00541-3.

27- Seif, Y., Arman, A. and Rahmanshahi, M., 2024. Multi-approach assessment for predicting submerged discharge reduction factor in porous broad-crested weirs. doi:10.21203/rs.3.rs-4110223/v1.

28- Shakeri Yousefi, S., Najarchi, M., Fuladipanah, M. and Rabani Bidgoli, M., 2025. Machine Learning models for High-Accuracy Prediction of Energy Dissipation Through Gabion Sills Downstream of Spillways. *Journal of New Approaches in Water Engineering and Environment*, 4(2), pp.91-106. doi:10.22034/nawee.2025.507107.1140.

29- Shariq, A., Hussain, A. and Ahmad, Z., 2022. Flow over gabion weir under free and submerged flow conditions. *Flow Measurement and Instrumentation*, 86, 102199. doi:10.1016/j.flowmeasinst.2022.102199.

30- Szczepanek, R., 2022. Daily streamflow forecasting in mountainous catchment using XGBoost, LightGBM and CatBoost. *Hydrology*, *9*(12), 226. doi:10.3390/hydrology9120226.

31- Tiwari, N.K., Luxmi, K.M. and Ranjan, S., 2023. Estimating gabion weir oxygen transfer with data mining. *Water Quality Research Journal*, *58*(1), pp.22–40. doi:10.2166/wqrj.2022.023.

EXTENDED ABSTRACT

Assessment of the Discharge Coefficient of Porous Weirs Using Hydraulic Experiments and Data-Driven Models

S. A. Esmaili[1], M. Heydari[2*] and S. Gohari[3]

1- Ph.D. Candidate, Department of Water Science and Engineering, Faculty of Agriculture, Bu-Ali Sina University, Hamedan, Iran
2- Corresponding Author, Associate Professor, Department of Water Science and Engineering, Faculty of Agriculture, Bu-Ali Sina University, Hamedan, Iran (mheydari@basu.ac.ir)*
3- Associate Professor, Department of Water Science and Engineering, Faculty of Agriculture, Bu-Ali Sina University, Hamedan, Iran

ARTICLE INFO

Article history:
Received:
Revised:
Accepted:
Available online:

Keywords: Gabion Weir, Flow Over, Machine Learning, XGBoost, CatBoost, GEP.

TO CITE THIS ARTICLE:

Abstract

Broad-crested rectangular gabion weirs are used as an effective means of regulating flow to an impoundment while conserving large amounts of energy and preventing adverse ecological impacts in hydraulic structures. This study aimed to develop various machine learning algorithms to estimate the coefficients of discharge for free-flowing water over the through-weir. Many of the machine learning regression algorithms were created from experimental data: RF, SR, SVR, XGBoost, and CatBoost, to find the best fit to model the coefficients of discharge from the dataset, with CatBoost and XGBoost providing the best overall performance. The total dataset was split so that 20% was used to test how well the models were performing, and the remaining 80% was used to train the models to provide the most accurate measure of performance for the models. The XGBoost model provided the best performance with an R2 of 0.96, NSE of 0.95, and RMSE of 0.029. The performance of the Simple Regression and Support Vector Regression models was less than desirable for each condition observed in comparison to that of the XGBoost model's performance. Finally, the 80:20 ratio of training-testing data sets was shown to be a large contributor to the predictive accuracy of the empirical relationship of the final equation produced by the GEP models.

Introduction

Among the various types of weirs, broad-crested gabion weirs have received considerable attention in hydraulic engineering research and design owing to their excellent structural stability and reliable discharge-measurement capability. The hydraulic behavior of these structures differs substantially from that of conventional impervious weirs due to their porous composition and the occurrence of internal flow through the gabion matrix. As porous hydraulic structures, gabion weirs permit flow both over the crest and through the body of the weir, resulting in more complex flow patterns and hydraulic interactions. In addition to their hydraulic functionality, gabion weirs offer several practical advantages, including relatively low construction costs, ease of installation, structural flexibility, and strong environmental compatibility. These attributes have contributed to their widespread use in river restoration, stream stabilization, erosion control, and ecological habitat enhancement projects.

Despite these advantages, the hydraulic performance of gabion weirs under varying flow conditions remains insufficiently understood and warrants further investigation. In particular, critical hydraulic parameters such as the discharge coefficient, flow distribution mechanisms, and

the influence of porosity on overall hydraulic behavior require comprehensive and systematic evaluation. Alomari et al. (2023) investigated the estimation of the discharge coefficient of rectangular broad-crested gabion weirs based on upstream flow depth and reported that the discharge coefficient increased by approximately 10.5% relative to conventional rigid weirs. In recent years, machine learning techniques have increasingly been applied to predict the hydraulic performance of gabion weirs. For instance, Tiwari et al. (2023) assessed several machine learning approaches for predicting the aeration efficiency of rectangular broad-crested gabion weirs and found that deep neural network models provided the highest predictive accuracy. Similarly, Hameed et al. (2025) employed hybrid reinforcement-based modeling approaches to estimate both the discharge coefficient and flow rate of rectangular broad-crested gabion weirs, demonstrating the effectiveness of hybrid intelligent models in predicting complex hydraulic parameters. Against this background, the primary objective of the present study is to develop and evaluate data-driven models for predicting the discharge coefficient (C_d) of rectangular broad-crested gabion weirs. The predictive capabilities of the proposed models are assessed against experimental laboratory measurements, and an explicit empirical formulation for C_d is subsequently derived using the Gene Expression Programming (GEP) algorithm to provide a practical and interpretable predictive tool for hydraulic engineering applications.

Methodology

The laboratory experiments were conducted in a rectangular glass-sided flume measuring 7.0 m in length, 0.60 m in width, and 0.50 m in depth at the Hydraulics Laboratory of the Hydraulic Structures Research Group, Ferdowsi University of Mashhad. A total of eight rectangular broad-crested gabion weir models were tested under a wide range of hydraulic and geometric conditions, including varying mean aggregate sizes, crest lengths, and discharge rates. To predict the free-flow discharge coefficient (C_d), five regression-based machine learning algorithms were implemented: Random Forest (RF), Stacked Regression (SR), Support Vector Regression (SVR), Extreme Gradient Boosting (XGBoost), and CatBoost. The predictive performance of the developed models was comprehensively assessed using five widely adopted statistical evaluation metrics, namely the Root Mean Square Error (RMSE), Mean Absolute Error (MAE), coefficient of determination (R^2), Scatter Index (SI), and Nash–Sutcliffe Efficiency (NSE).

Results and Discussion

The discharge coefficient (C_d) was selected as the target variable for model development. The compiled dataset was primarily divided into training (80%) and testing (20%) subsets. To further evaluate model robustness and predictive sensitivity, additional data-partitioning scenarios with training proportions of 70%, 60%, and 50% were also considered. Based on the statistical performance metrics, the XGBoost model developed under the 80:20 training-to-testing ratio exhibited the best overall performance, achieving a coefficient of determination (R^2) of 0.960, a root mean square error (RMSE) of 0.065, a mean absolute error (MAE) of 0.041, and a scatter index (SI) of 0.076. A reduction in the size of the training dataset resulted in a noticeable decline in predictive performance, with R^2 values decreasing to 0.827, 0.833, and 0.812 for the 70%, 60%, and 50% training scenarios, respectively. Under the baseline 80:20 data-partitioning framework, the CatBoost model ranked second and demonstrated comparable predictive capability, achieving R^2, RMSE, MAE, and SI values of 0.962, 0.042, 0.026, and 0.050, respectively. In contrast, the Random Forest (RF), Stacked Regression (SR), and Support Vector Regression (SVR) models exhibited lower predictive accuracy under the same data allocation, with R^2 values of 0.898, 0.899, and 0.882, respectively. Their performance deteriorated further when the training proportion was reduced to 50%, yielding R^2 values of 0.690, 0.712, and 0.563, respectively.

These results clearly indicate that the predictive capability of machine learning models is strongly influenced by the adopted data-partitioning strategy. The superior performance of the gradient-boosting algorithms, particularly XGBoost and CatBoost, can be attributed to their enhanced ability to capture the highly nonlinear relationships between weir geometry, hydraulic parameters, and flow characteristics, which are essential for accurately modeling discharge behavior over porous gabion weirs (see Fig.1).

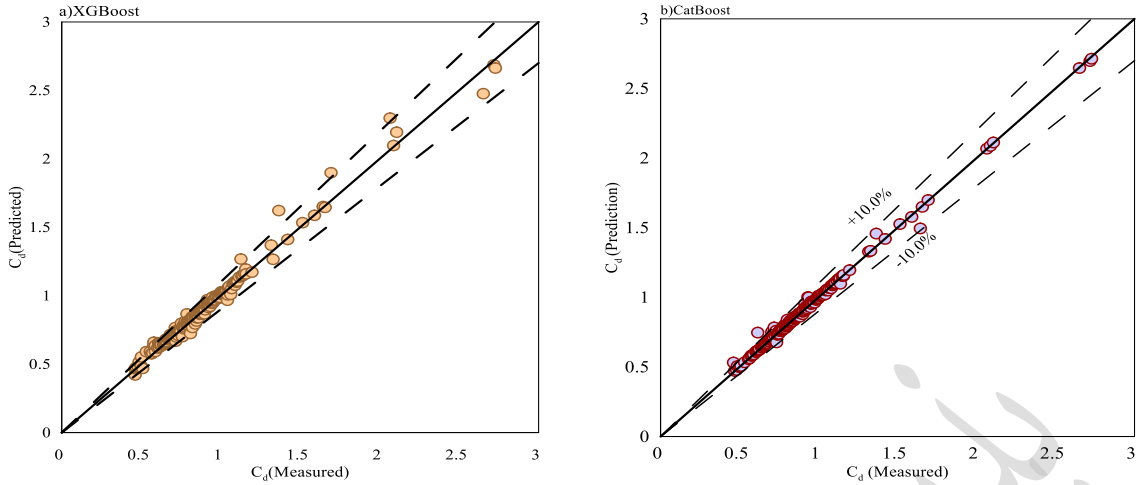


Fig. 1- Results of testing of the considered models with the (80-20) split in C_d prediction

Fig. (2) presents the Taylor diagram, which provides a graphical validation of the statistical performance metrics obtained for the investigated machine learning models. As shown, the comparative assessment in Taylor space is in close agreement with the trends identified from the MAE and RMSE analyses. The XGBoost model, characterized by a correlation coefficient greater than 0.99, a normalized standard deviation of 0.32, and the lowest root-mean-square error, exhibited the highest predictive performance, demonstrating an exceptional capability to capture the complex hydraulic flow processes governing gabion-weir behavior. This model not only achieved the lowest MAE and RMSE values but also displayed superior robustness and generalization ability by closely reproducing the observed variability and attaining the highest correlation coefficient in the Taylor diagram. Relative to the Stacked Regression (SR) model, these results highlight the outstanding predictive accuracy and computational efficiency of XGBoost for estimating the discharge coefficient (C_d) of gabion weirs.

In contrast, the CatBoost model showed a tendency to overestimate the variance, whereas the Random Forest (RF) model ranked lower due to its comparatively weaker correlation and moderate predictive accuracy. The SR model exhibited the poorest performance among the investigated approaches, failing to adequately reproduce the observed variability and producing relatively large error distributions, thereby limiting its applicability for accurate discharge-coefficient prediction in this hydraulic context.

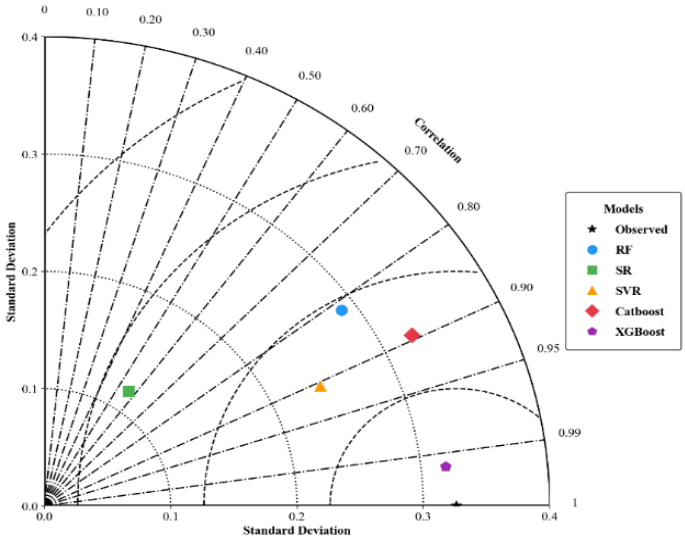


Fig. 2- Taylor diagram illustrating the performance of machine learning models in predicting C_d

Following the development of the black-box predictive models, Gene Expression Programming (GEP) was employed using GeneXproTools software to derive explicit empirical equations based on the governing dimensionless parameters. The GEP models developed for the entire dataset (Eq. 1), as well as for the 80% (Eq. 2) and 70% (Eq. 3) training-data scenarios, demonstrated satisfactory predictive performance. For the 80% training scenario, the model achieved R^2, RMSE, and MAE values of 0.955, 0.106, and 0.055, respectively, whereas the corresponding values for the 70% training scenario were 0.942, 0.084, and 0.061. Finally, an explicit predictive equation was established by training the GEP algorithm using the complete dataset, thereby providing a generalized empirical formulation for estimating the discharge coefficient.

$$C_d = \frac{Fr^3 P}{H} - 1.112\frac{Fr^2 P}{H} + \frac{H^2}{LP} - \frac{H^3}{L^3} + \frac{1}{0.092(\frac{H}{L}) + \frac{H}{PFr}} \quad (1)$$

$$C_d = \frac{0.846\, FrLP}{H(L+P)} + \frac{HFr^2}{L} + \frac{H^2}{LP} - 0.665\frac{H}{L} + \sqrt{\frac{HFr}{L}} + \frac{HLFr}{P(H-2L)} \quad (2)$$

$$C_d = \frac{\frac{H}{L}}{Fr(\frac{P}{H} + 5.898) - \frac{H}{P}} + \frac{H^2(L-P)}{12.109L^2P} - 0.177\frac{H}{L} + Fr + 0.1805\, Fr\left(\frac{P}{H}\right)^{1.5} \quad (3)$$

Conclusions

Rectangular broad-crested gabion weirs are extensively utilized in open channels, natural rivers, and small- to medium-scale hydraulic structures owing to their reliable discharge measurement capability, effective flow regulation, favorable energy dissipation characteristics, structural stability, and environmental compatibility. In the present study, the predictive performance and robustness of five regression-based machine learning models, namely Random Forest (RF), Stacked Regression (SR), Support Vector Regression (SVR), and the advanced gradient-boosting algorithms XGBoost and CatBoost, were systematically evaluated and compared under four data-partitioning scenarios with training proportions of 50%, 60%, 70%, and 80% to predict the discharge coefficient (C_d) of gabion weirs.

The results demonstrated that the size of the training dataset exerts a substantial influence on model generalization capability, with the 80:20 training-to-testing ratio providing the most effective framework for capturing the inherent nonlinear hydraulic flow behavior. Among the investigated models, the gradient-boosting algorithms consistently outperformed the conventional machine learning approaches in estimating the discharge coefficient. Notably, the XGBoost and CatBoost models developed under the 80:20 data-partitioning scenario exhibited the highest predictive accuracy and stability, with no evidence of overfitting, achieving the highest coefficients of determination and superior overall predictive performance. Furthermore, the development of explicit empirical equations using Gene Expression Programming (GEP) confirmed that the data-partitioning strategy plays a critical role in determining the accuracy, reliability, and predictive robustness of the resulting governing formulations.

References

1- Alomari, N.K., Altalib, A.N. and Al-Janabi, A.M.S., 2023. Discharge estimation using brink depth over a trapezoidal-shaped weir. *Flow Measurement and Instrumentation*, 94, 102454. doi:10.1016/j.flowmeasinst.2023.102454.

2- Hameed, M.M., Alomar, M.K., Razali, S.F.M. and Salem, A., 2025. Integrated approach of extreme learning machines and locally weighted linear regression for improved discharge coefficient prediction. *Scientific Reports*, 15(1), 21761. doi:10.1038/s41598-025-03812-z.

3- Tiwari, N.K., Luxmi, K.M. and Ranjan, S., 2023. Estimating gabion weir oxygen transfer with data mining. *Water Quality Research Journal*, 58(1), pp.22–40. doi:10.2166/wqrj.2022.023.

پذیرش نهایی- قبل از چاپ