

Sensitivity analysis and discharge coefficient estimation in semicircular weirs using machine learning methods

Amal Savaedi ¹, Mohammadreza Zayeri ^{2✉}, Mehdi Ghomeshi ³, and Mehdi Daryaei ⁴

1. Department of Hydraulic Structures, Faculty of Water Engineering and Environmental Science, Shahid Chamran University of Ahvaz, Ahvaz, Iran. E-mail: amalsavaedi98@gmail.com
2. Corresponding author, Department of Hydraulic Structures, Faculty of Water Engineering and Environmental Science, Shahid Chamran University of Ahvaz, Ahvaz, Iran. E-mail: m.zayeri@scu.ac.ir
3. Department of Hydraulic Structures, Faculty of Water Engineering and Environmental Science, Shahid Chamran University of Ahvaz, Ahvaz, Iran. E-mail: ghomeshi@scu.ac.ir
4. Department of Hydraulic Structures, Faculty of Water Engineering and Environmental Science, Shahid Chamran University of Ahvaz, Ahvaz, Iran. E-mail: m.daryaei@scu.ac.ir

Article Info.	ABSTRACT
Article type: Research Article	Objective This study aims to develop and evaluate two machine learning models, namely Extreme Learning Machine (ELM) and eXtreme Gradient Boosting (XGBoost), incorporating Bayesian Optimization and cross-validation techniques, for the prediction of the discharge coefficient (Cd) of semicircular side weirs.
Article history: Received 05 June 2025 Received in revised form 09 October 2025 Accepted 16 December 2025 Available online 22 June 2026	Method: The modeling process is based on a dataset comprising 233 experimental observations sourced from two independent studies, encompassing both geometric and hydraulic parameters. Bayesian Optimization was implemented using five-fold cross-validation on the training set. For each parameter combination, the root mean square error (RMSE) was computed, and once the objective function (RMSE) reached a satisfactory threshold, the corresponding parameter set was selected as the final model structure. The optimized model was subsequently validated using the test dataset, and its predictive performance was compared against baseline models.
Keywords: discharge coefficient, extreme learning machine, XGBoost, sensitivity analysis, Harmonic weir.	Results: The results indicated that the XGBoost model achieved higher predictive accuracy with a coefficient of determination (R^2) of 0.99 for the training set and 0.92 for the test set, outperforming the ELM model, which yielded R^2 values of 0.91 and 0.879 for the respective datasets. Sensitivity analysis using Sobol, Morris, Entropy-based, and SHAP methods consistently identified the HT/P ratio (hydraulic head to weir height) as the most influential input parameter in predicting Cd. Specifically, Sobol analysis revealed that nearly the entire variance in the model output could be attributed to this single parameter, with negligible interaction effects from other variables.
	Conclusions: All sensitivity analysis techniques confirmed the robustness and reliability of the developed models. The integration of machine learning algorithms with Bayesian Optimization proves to be an effective and powerful framework for flow modeling in hydraulic systems.
Cite this article: Savaedi, A., Zayeri, M.R., Ghomeshi, M., & Daryaei, M. (2026). Sensitivity Analysis and Discharge Coefficient Estimation in Semicircular Weirs Using Machine Learning Methods. <i>Advanced Technologies in Water Efficiency</i> , 6(2), 1-23. https://doi.org/10.22126/atwe.2025.12271.1170	
 © The Author(s) Publisher: Razi University. https://doi.org/10.22126/atwe.2025.12271.1170	

Introduction

Dams, as critical hydraulic structures, consist of key components including the dam body, drainage systems, and most notably, spillways, which serve as primary outlets for flood discharge. When the capacity to release floodwaters is limited, the choice of spillway type becomes essential to optimize both discharge efficiency and construction costs. Among the various options, labyrinth spillways—with non-linear crest geometries—have proven to be highly effective in enhancing discharge capacity while managing upstream water levels during flood events. By significantly increasing the effective crest length over a fixed width, these geometries, such as triangular, trapezoidal, semicircular, and polygonal plans, outperform traditional linear (rectangular) spillways.

Recent research efforts have focused on understanding the influence of hydraulic and geometric parameters on discharge coefficient (C_d) and velocity distribution in these structures. Numerous machine learning approaches have been employed to predict C_d , with Support Vector Machines (SVM), Artificial Neural Networks (ANN), and Extreme Learning Machines (ELM) demonstrating promising accuracy and adaptability. In particular, studies have shown the superior predictive performance of hybrid models such as SVR-ISADE, GASVM, and ANFIS-FFA compared to classical regression methods. The semicircular labyrinth weir (SCLW) is of particular interest due to its high energy dissipation and superior discharge capabilities—achieving up to 4.5 times the efficiency of linear weirs. Sensitivity analyses across multiple studies have highlighted the hydraulic head-to-weir height ratio (H/P) as the most influential parameter affecting C_d . Despite the complex physical behavior of SCLWs, data-driven and AI-based models have shown remarkable precision in C_d estimation.

This study develops novel predictive models for the discharge coefficient of harmonic semicircular labyrinth weirs using XGBoost and Extreme Learning Machines (ELM). It further quantifies the effect of non-dimensional variables on C_d and aims to bridge existing research gaps by integrating and analyzing experimental data from two previous investigations.

Method

In this study, a comprehensive modeling framework is developed to predict the discharge coefficient (C_d) of semicircular labyrinth weirs using advanced machine learning techniques and sensitivity analysis methods. The first stage involves dimensional analysis and data compilation. The discharge coefficient is calculated from physical experiments using dimensional parameters such as the number of weir cycles (N), radius of curvature (R), crest height (P), and channel width (B). The effects of fluid density, viscosity, and surface tension are deemed negligible due to turbulent flow and sufficient hydraulic head, following guidelines by Rao & Shakla (1971). Using dimensional analysis, the governing relationship is simplified to a functional form involving non-dimensional parameters: H/P , L/P , B/P , R/P , and N . A dataset of 233 experimental results was compiled from prior studies (Namazi & Mozaffari, 2023; Yildiz et al., 2024), covering multiple configurations of weir height and the number of cycles. 75% of the data was used for training and 25% for testing the machine learning models. Descriptive statistics and histograms of input variables confirm weak linear correlations with C_d , justifying the use of nonlinear models such as Extreme Learning

Machine (ELM) and XGBoost. The ELM model, comprising an input layer, hidden neurons with random weights, and a single output node for C_d , was trained with Bayesian Optimization to identify optimal structure (number of neurons and activation function). Conversely, XGBoost builds an ensemble of regression trees sequentially, where each tree reduces the residuals of the previous one. The model's objective function incorporates both prediction loss and a regularization term to avoid overfitting. For feature importance assessment, four global sensitivity analysis techniques were implemented. The Morris method computes mean absolute elementary effects to detect influential variables. Sobol's method decomposes output variance into first-order and total sensitivity indices. Entropy-based analysis calculates mutual information to evaluate nonlinear dependencies. Lastly, the SHAP method applies cooperative game theory to quantify each input's contribution to model prediction.

Results

Results consistently identify HT/P (relative total head to crest height) as the most influential parameter across all sensitivity methods. Sobol indices indicate that nearly all output variance is explained by HT/P alone. The entropy method assigns HT/P a normalized importance of 0.87, and the Morris μ^* score further supports its dominant role. This multi-method framework demonstrates the potential of hybrid machine learning and dimensional analysis to model complex hydraulic behaviors in labyrinth weirs, offering reliable predictive tools for engineering design.

Conclusions

Accurate prediction of the discharge coefficient (C_d) in semicircular harmonic weirs is crucial for the optimal design and performance of hydraulic systems. Due to the complex relationship between C_d and its influencing parameters, empirical equations face limitations. In this study, two hybrid machine learning models—ELM and XGBoost—were developed and optimized using Bayesian Optimization to model C_d . Feature importance analysis was conducted using four sensitivity methods to determine the relative influence of input variables. Among the approaches, combining XGBoost with SHAP-based sampling demonstrated the most stable and precise C_d predictions, offering narrower uncertainty bands and greater resistance to noise. However, despite XGBoost's advanced structure and tunable parameters, ELM outperformed it in terms of accuracy. The Taylor diagram showed greater distance of XGBoost from the reference point compared to ELM in both training and testing sets, highlighting its lower prediction precision. These results can guide model selection and sensitivity analysis in future uncertainty-based studies.

Author Contributions

All authors contributed equally to the conceptualization of the article and writing of the original and subsequent drafts.

Data Availability Statement

Data available on request from the authors.

Acknowledgements

The authors would like to thank all participants in the present study.

Ethical Considerations

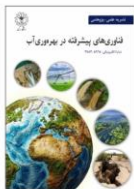
The authors avoided data fabrication, falsification, plagiarism, and misconduct.

Funding

The authors are grateful for the financial support of the Research Council of Shahid Chamran University of Ahvaz (GN: SCU.WH1403.43525).

Conflict of Interest

The authors declare no conflict of interest



آنالیز حساسیت و برآورد ضریب دبی در سرریزهای نیم‌دایره با استفاده از روش‌های یادگیری

ماشین

امل سواعدی^۱، محمدرضا زایری^{۲*}، مهدی دریایی^۳، و مهدی قمشی^۴

۱. گروه سازه‌های آبی، دانشکده مهندسی آب و محیط زیست، دانشگاه شهید چمران اهواز، اهواز، ایران. رایانامه: amalsavaedi98@gmail.com

۲. نویسنده مسئول، گروه سازه‌های آبی، دانشکده مهندسی آب و محیط زیست، دانشگاه شهید چمران اهواز، اهواز، ایران. رایانامه: m.zayri@scu.ac.ir

۳. گروه سازه‌های آبی، دانشکده مهندسی آب و محیط زیست، دانشگاه شهید چمران اهواز، اهواز، ایران. رایانامه: ghomeshi@scu.ac.ir

۴. گروه سازه‌های آبی، دانشکده مهندسی آب و محیط زیست، دانشگاه شهید چمران اهواز، اهواز، ایران. رایانامه: m.daryaei@scu.ac.ir

در باره مقاله	چکیده
نوع مقاله: مقاله پژوهشی	هدف: در این پژوهش، دو مدل یادگیری ماشین شامل ELM و XGBoost با رویکرد بهینه‌سازی بیزی و اعتبارسنجی با هدف پیش‌بینی ضریب دبی سرریزهای نیم‌دایره‌ای توسعه داده شدند.
تاریخ دریافت: ۱۴۰۴/۰۳/۱۵ تاریخ بازنگری: ۱۴۰۴/۰۷/۱۷ تاریخ پذیرش: ۱۴۰۴/۰۹/۲۵ تاریخ انتشار: ۱۴۰۵/۰۴/۰۱	روش پژوهش: در پژوهش حاضر، با اتکا به مجموعه داده‌های متشکل از ۲۳۳ داده آزمایشگاهی از دو منبع پژوهشی، شامل پارامترهای هندسی تحت سه ارتفاع متفاوت برای سرریز (معادل ۱۵، ۲۰، ۳۰ و ۴۰ سانتی‌متر) و با چهار تعداد مختلف از سیکل هارمونیک و همچنین متغیرهای هیدرولیکی، به مدل‌سازی ضریب دبی پرداخته شده است. شش پارامتر کلیدی در مدل XGBoost و دو پارامتر کلیدی در مدل ELM با استفاده از روش بیزی بهینه‌سازی شده است.
کلیدواژه‌ها: ضریب دبی، مدل یادگیری افراطی (ELM)، الگوریتم XGBoost، تحلیل حساسیت، سرریز هارمونیک.	یافته‌ها: نتایج نشان داد که مدل XGBoost با ضریب تعیین (R^2) برابر با ۰/۹۹ در آموزش و ۰/۹۲ در آزمون عملکرد دقیق‌تری نسبت به مدل ELM با R^2 معادل ۰/۹۱ و ۰/۸۷۹ در مجموعه‌های مشابه داشته است. همچنین تحلیل حساسیت با بهره‌گیری از روش‌های سوپول، مورس، آنتروپی و SHAP، پارامتر HT/P یعنی نسبت هد هیدرولیکی به ارتفاع سرریز را به‌عنوان مؤثرترین ورودی در پیش‌بینی ضریب دبی شناسایی کرد. همچنین تمام واریانس پاسخ مدل در روش آنالیز حساسیت سوپول را می‌توان تنها با تغییر در این پارامتر توجیه کرد و تأثیر متقابل آن با سایر پارامترها بسیار ناچیز است.
	نتیجه‌گیری: نتایج حاصل از تمامی روش‌های تحلیل حساسیت بر پایایی و دقت مدل‌های توسعه‌یافته تأکید داشته و نشان می‌دهند که رویکرد ترکیبی یادگیری ماشین و بهینه‌سازی بیزی می‌تواند ابزاری توانمند برای مدل‌سازی جریان در پیش‌بینی ضریب دبی سرریزهای نیم‌دایره‌ای محسوب شود.

استناد: سواعدی، امل؛ زایری، محمدرضا؛ قمشی، مهدی؛ و دریایی، مهدی. (۱۴۰۵). آنالیز حساسیت و برآورد ضریب دبی در سرریزهای نیم‌دایره با استفاده از

روش‌های یادگیری ماشین. *فناوری‌های پیشرفته در بهره‌وری آب*، ۶ (۲)، ۱-۲۳.

<http://doi.org/10.22126/atwe.2025.12271.1170>



© نویسندگان

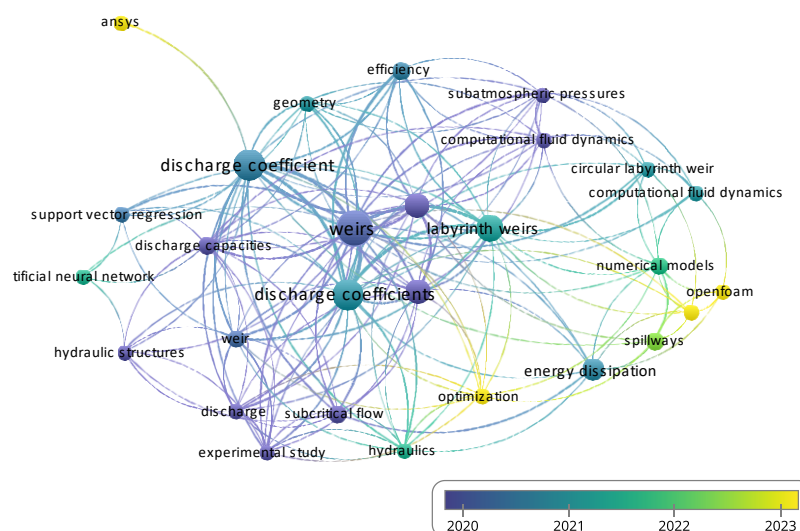
شر: دانشگاه رازی.

مقدمه

سد ها از ارکان اصلی چون بدنه سد، سازه تخلیه کننده سیلاب یا همان سرریز، و سیستم های زهکشی تشکیل یافته اند. چنانچه توان رهاسازی سیلاب از مخزن سد محدود باشد، انتخاب نوع سرریز، در واقع، تلاشی است در راستای بهینه سازی فرآیندی که در آن، دستیابی به بیشترین ظرفیت عبور جریان از سرریز، توأم با حداقل هزینه های اجرایی، مد نظر قرار می گیرد. در این میان، طول و هندسه تاج سرریز دو پارامتر کلیدی، نقش تعیین کننده ای در میزان دبی قابل عبور از آن ایفا می کنند. برای مواجهه با این چالش و به منظور ارتقاء ظرفیت تخلیه، بهره گیری از سرریزهای با پلان غیرمستقیم، موسوم به سرریزهای کنگره ای، به عنوان یک راهکار مؤثر مطرح می گردد. این نوع خاص از سرریزها، نه تنها توان عبور جریان را به طرز چشمگیری بهبود می بخشد، بلکه در مهار افزایش سطح آب در مخزن به هنگام وقوع سیلاب نیز کارآمد ظاهر می شوند. افزون بر این، چنین ساختارهایی در نقش مستهلک کننده های کارآمد انرژی جنبشی جریان، بهبود فرایند هوادهی آب و به طور کلی، سازه هایی برای افت انرژی، کاربردهای قابل توجهی می یابند (قادری و همکاران^۱، ۲۰۲۴؛ حسین و جلیل^۲، ۲۰۲۴؛ اکینگلاری^۳، ۲۰۲۴). سرریزهای کنگره ای در اشکال و طرح های گوناگونی، نظیر پلان های مثلثی، دوزنقه ای، دایره ای و یا چندضلعی، موجود می باشند. ویژگی برجسته و امتیاز کلیدی این نوع از سرریزها، در مقایسه با سرریزهای خطی (مستطیلی)، در توانایی آنها برای افزایش قابل ملاحظه طول مؤثر مسیر جریان بر روی عرض ثابت، می باشد. پژوهشگران، به طور پیوسته، در پی یافتن و شناسایی طرحی بهینه برای این سازه ها هستند؛ طرحی که در عین تضمین عملکردی در سطح بالا، بتواند هزینه های اجرایی را نیز به نحو محسوسی کاهش دهد.

ادبیات موضوع و پیشینه پژوهش

برای بررسی بیشتر در این حوزه، شکل (۱) به نمایش و تحلیل کلیدواژه هایی می پردازد که پژوهشگران پیشین در مطالعات مرتبط با ضریب دبی جریان در سرریزها، بر اساس داده های پایگاه استنادی اسکوپوس، بیشترین کاربرد را برای آنها در نظر دارند. چنانکه از این بررسی استنباط می شود، پیوند روش های میان تکنیک های بهینه سازی و روش های دینامیک سیالات محاسباتی (CFD) با طراحی و تحلیل سرریزهای کنگره ای، خود گواهی بر تلاش های مستمر جهت بهبود و ارتقاء عملکرد این سازه ها در طول پنج سال اخیر است.



شکل ۱. واژگان کلیدی در تحقیقات گزارش شده در مدل سازی هیدرولیکی سرریزهای کنگره ای

¹ Ghaderi et al

² Hussein & Jalil

³ Ikinciogullari

تلاش‌های پژوهشی قابل‌توجهی در جهت چگونگی تأثیرگذاری پارامترهای هیدرولیکی و مشخصه‌های هندسی بر مؤلفه‌های کلیدی جریان، نظیر ضریب دبی و توزیع سرعت، در این دسته از سازه‌های هیدرولیکی انجام گردیده است (بوریسوا و بلیکوف^۱، ۲۰۲۳؛ سلیمانی^۲، ۲۰۲۲؛ زارع^۳، ۲۰۲۳). برقی و پروانه^۴ (۲۰۱۱) خاطرنشان کردند که عملکرد سرریز جانبی مایل، به لحاظ کارایی، برتری چشمگیری معادل ۲.۳۳ برابر نسبت به سرریز جانبی معمولی به نمایش می‌گذارد. کیسی و همکاران^۵ (۲۰۱۲) توانمندی شبکه‌های عصبی با توابع پایه شعاعی (RBNN) و شبکه‌های عصبی رگرسیون عمومی (GRNN) در پیش‌بینی دبی عبوری از سرریزهای جانبی با مقطع مثلثی، مورد ارزیابی قرار گرفت. نتایج حاصله، گواهی بر این است که این الگوهای محاسباتی، در بازنمایی و شناسایی روابط غیرخطی میان مشخصه‌های هندسی سازه و میزان جریان عبوری، از قابلیت و اثربخشی خوبی برخوردارند. سیمسک و همکاران^۶ (۲۰۲۳)، به‌منظور پیش‌بینی ضریب دبی (Cd) در سرریزهایی با مقطع دوزنقه‌ای، از شبکه‌های عصبی مصنوعی (ANN) مورد استفاده قرار دادند. یافته‌های پژوهش ایشان آشکار ساخت که پارامتر بی‌بعد فرود، نقشی تعیین‌کننده در بهبود چشمگیر عملکرد مدل‌ها جهت برآورد مقادیر Cd ایفا می‌کند؛ به‌گونه‌ای که می‌توان ادعا کرد که رویکرد مبتنی بر شبکه‌های عصبی مصنوعی، در مقایسه با سایر رویکردهای مرسوم، در تعیین ضریب دبی، از کارایی و دقت بهتری برخوردار است. بیلهان و همکاران^۷ (۲۰۱۸) به مقایسه و تطبیق داده‌های آزمایشگاهی به‌دست‌آمده از مطالعه بر روی سرریزهای نیم‌دایره‌ای با نتایج حاصل از شبیه‌سازی‌های عددی که با بهره‌گیری از نرم‌افزار Flow-3D انجام پذیرفته بود، مقایسه کردند. مدل‌های عددی توسعه‌یافته توسط ایشان، انطباق قابل ملاحظه‌ای، تا سطح ۹۶ درصد، با داده‌های تجربی به نمایش گذاشتند؛ که نشان از، کارآمدی رویکردهای مدل‌سازی عددی را در تخمین دقیق دبی عبوری از سرریز می‌باشد. افزون بر این، یافته‌های پژوهشان نشان از آن دارد که عملکرد ضریب جریان، در تابعیت از نسبت هد هیدرولیکی کل به ارتفاع سرریز (Ht/P)، برای سرریزهای با مقطع نیم‌دایره‌ای، از الگوی مشابهی با سرریزهای دوزنقه‌ای که دارای زوایای دیواره جانبی فراتر از ۳۵ درجه هستند، پیروی می‌کند. سنگ سفیدی و همکاران^۸ (۲۰۱۸)، ضمن انجام یک بررسی آزمایشگاهی بر سرریزهای کنگره‌ای با هندسه دایره‌ای شکل دنداندار، به این یافته دست یافتند که توانایی تخلیه دبی در این نوع از سرریزها، در مقایسه با سازه‌های خطی، افزایشی معادل چهار و نیم برابر را به نمایش می‌گذارد. دووتا و همکاران^۹ (۲۰۲۰)، به‌منظور ارزیابی توانمندی مدل‌های رگرسیون خطی چندگانه (MLR)، ماشین‌های بردار پشتیبان (SVM) و شبکه‌های عصبی مصنوعی (ANN) در پیش‌بینی دبی برای گونه‌های مختلف سرریز (اعم از سرریزهای استاندارد و نیم‌دایره‌ای)، پژوهشی را به انجام رساندند. با استناد به معیارهای آماری چون R^2 ، APE و CE، به این جمع‌بندی دست یافتند که الگوریتم ماشین بردار پشتیبان (SVM)، در مقایسه با سایر رویکردها، عملکرد بهتری از خود نشان داده است. امری که گواهی بر قابلیت برتر آن در شناسایی و مدل‌سازی روابط پیچیده میان مؤلفه‌های جریان و میزان دبی عبوری، در خصوص این اشکال خاص از سازه‌های سرریز، است. شفیعی و همکاران^{۱۰} (۲۰۲۰)، با اتکا به یک رویکرد مبتنی بر هوش مصنوعی (AI)، به برآورد ضریب دبی در تاج سرریزها پرداختند. یافته‌های پژوهش ایشان نشانگر آن بود که مدل ORELM، حتی با وجود چالش‌هایی

¹ Borisova & Belikov

² Suleimany et al

³ Zare et al

⁴ Borghei & Parvaneh

⁵ Kisi et al

⁶ Simsek et al

⁷ Bilhan et al

⁸ Sangsefidi et al

⁹ Dutta et al

¹⁰ Shafiei et al

نظیر عدم وجود یک رابطه تحلیلی صریح، در شبیه سازی ضریب دبی، عملکردی مطلوب و قابل قبول از خود به نمایش گذاشته است. شفیی و همکاران (۲۰۲۰) ایشان، با تلفیق سیستم استنتاج عصبی-فازی تطبیقی (ANFIS) و الگوریتم جلبک آتشین (FFA)، به برآورد ضریب دبی در تاج سرریزهای مثلثی پرداختند. یافته های این پژوهش، مبین آن بود که مدل ترکیبی ANFIS-FFA در عرصه پیش بینی ضریب دبی برای سرریزهای مثلثی شکل، عملکردی برتر و کارایی بهتری را در مقایسه با مدل استاندارد ANFIS به نمایش می گذارد. بنکداری و همکاران^۱ (۲۰۲۰)، با بهره گیری از تکنیک برنامه نویسی بیانی ژن (GEP)، به برآورد ضریب دبی در تاج سرریزهای مثلثی پرداختند؛ نتایج حاصل از این پژوهش، گواهی بر آن است که روش GEP در پیش بینی ضریب جریان برای این نوع از سرریزها، به نتایجی مطلوب انجام گردیده است. امامی و همکاران^۲ (۲۰۲۱)، در راستای برآورد ضریب تخلیه در تاج سرریزهایی با هندسه W-شکل، به توسعه یک الگوریتم تکامل تفاضلی خودتطبیق بهبودیافته (ISADE) در هم افزایی با رگرسیون بردار پشتیبان (SVR) که از آن با عنوان SVR_ISADE یاد می شود، پرداخته و نتایج حاصل، گواهی بر عملکرد برتری رویکرد SVR_ISADE در پیش بینی ضریب دبی برای این دسته از سرریزها بود؛ چنانکه دستیابی به مقادیر R^2 برابر با ۰/۹۸۲، MAPE معادل ۰/۶۱۲، RMSE برابر با ۰/۰۰۶ و δ معادل ۰/۸۴۳ که نشان از عملکرد برتر این مدل است. صفررضوی زاده و همکاران^۳ (۲۰۲۱)، در یک بررسی آزمایشگاهی، به بررسی تأثیر پارامترهایی چون نسبت هد هیدرولیکی در بالادست به ارتفاع سازه (H_T/P)، نسبت طول سرریز به ارتفاع (L/P)، نسبت شعاع انحنا به ارتفاع (R/P) و تعداد سیکل های سرریز (N) بر ضریب دبی در سرریزهای کنگره ای با پلان نیم دایره و سینوسی پرداختند. نتایج ایشان نشان می دهد که در بازه نسبت های ($H_T/P < 0.35$)، ضریب دبی برای هر دو نوع سرریز کنگره ای ابتدا روندی صعودی و سپس نزولی را تجربه می کند. علاوه بر این، مشخص گردید که ظرفیت عبور جریان در این سرریزهای کنگره ای، در مقایسه با سرریزهای خطی، به میزان ۳۳ درصد برتری دارد. سواعدی و همکاران^۴ (۲۰۲۴)، با استفاده از سری داده آزمایشگاهی، به بررسی تأثیر پارامترهای هندسی و هیدرولیکی بر ضریب دبی سرریز کنگره ای نیم دایره ای پرداختند. آن ها با به کارگیری سه مدل یادگیری ماشین شامل رگرسیون بردار پشتیبان (SVR)، مدل MARS و مدل جنگل تصادفی (RF)، مدل MARS را به عنوان دقیق ترین روش در پیش بینی ضریب دبی معرفی کردند و عمق نسبی (H_T/P) را مؤثرترین پارامتر دانستند. سلیمانی و همکاران^۵ (۲۰۲۲)، با به کارگیری روش های رگرسیون خطی چندگانه (MLR) و رگرسیون غیرخطی، در پی تدوین یک معادله تجربی برای تخمین دبی سرریزهای با مقطع نیم دایره ای برآمدند. مدل های ارائه شده توسط ایشان، با دستیابی به ضرایب تعیین (R^2) به مقادیر ۰/۸۴۹۸ و ۰/۸۵۸۴، انطباق قابل قبولی را با داده های آزمایشگاهی به اثبات رساندند. حقیبن و همکاران^۶ (۲۰۱۸)، با ارائه یک رویکرد داده محور، به تخمین ضریب دبی (Cd) در سرریزهای پله ای پرداختند. مدل بهینه سازی شده توسط ایشان، بهبود چشمگیر ۸۶/۱۳ درصدی در شاخص عملکرد را به ارمغان آورد؛ این دستاورد، تایید بر پتانسیل قابل توجه روش های مبتنی بر داده در افزایش دقت پیش بینی های مربوط به دبی است. لی و همکاران^۷ (۲۰۲۴)، با بهره گیری از ماشین های بردار پشتیبان (SVM) در تلفیق با سه الگوریتم بهینه سازی، به توسعه مدلی جهت پیش بینی ضریب دبی (Cd) در سرریزهای کناری با مقطع نیم دایره (SCSWs) پرداختند. یافته های ایشان تاکید بر آن است که رویکرد مبتنی بر ماشین بردار پشتیبان و الگوریتم ژنتیک (GASVM)، از دقت برتری در پیش بینی و قابلیت

¹ Bonakdari et al² Emami et al³ Safarrazavi et al⁴ Savaedi et al⁵ Suleimany et al⁶ Suleimany et al⁷ Li et al

تعمیم‌پذیری مطلوبی برخوردار است؛ به گونه‌ای که میزان خطای متوسط و بیشینه خطا به ترتیب بالغ بر ۰/۰۸ و ۰/۴۷ درصد گردید؛ این مقادیر، در مقایسه با الگوهای تجربی رایج، به ترتیب کاهشی در حدود ۷۲/۹۵ و ۸۶/۶۰ درصد را به نمایش می‌گذارند.

نظر به ماهیت فیزیکی پیچیده سازه‌های سرریز کنگره‌ای با پلان نیم‌دایره، استفاده از الگوهای مبتنی بر هوش مصنوعی که از دقت و پایداری بالایی در جهت تخمین ضریب دبی، با بررسی متون علمی برخوردار است می‌تواند الگویی دقیق و جامع برای پیش‌بینی ضریب تخلیه (Cd) در این گونه خاص از سازه‌ها توسعه دهد. از این رو، پژوهش حاضر به ارائه مدل‌هایی برای پیش‌بینی ضریب دبی (Cd) سرریزهای با پلان نیم‌دایره هارمونیک، با بهره‌گیری از شیوه‌هایی نظیر الگوریتم XGBoost و الگوی ماشین یادگیری افراطی^۱ (ELM)، می‌پردازد. افزون بر این، در این تحقیق، اثرگذاری متغیرهای بی‌بعد بر ضریب دبی به صورت کمی با استفاده از روش‌های آنالیز حساسیت سوبول، موریس، آنترپوی و SHAP تبیین گردیده و با تجمیع و تحلیل داده‌های آزمایشگاهی حاصل از دو پژوهش پیشین، تلاش شده است تا خلأ مطالعاتی در زمینه مدل‌سازی ضریب دبی مرتفع گردد.

روش پژوهش

۱. آنالیز ابعادی و جمع آوری داده‌ها

ضریب دبی (Cd) برای سرریزهای کنگره‌ای با پلان نیم‌دایره، به واسطه معادله (۱) محاسبه می‌گردد. در این معادله، Q بیانگر دبی جریان، L معرف طول سرریز، g نماد شتاب ثقل، و H_T نشان‌دهنده هد هیدرولیکی کل است:

$$Q = \frac{2}{3} C_d L \sqrt{2gH_T^{1.5}} \quad (1)$$

پارامترهای مؤثر در تعیین ضریب دبی (Cd) برای سرریزهای کنگره‌ای با پلان نیم‌دایره را می‌توان به صورت زیر در نظر گرفت:

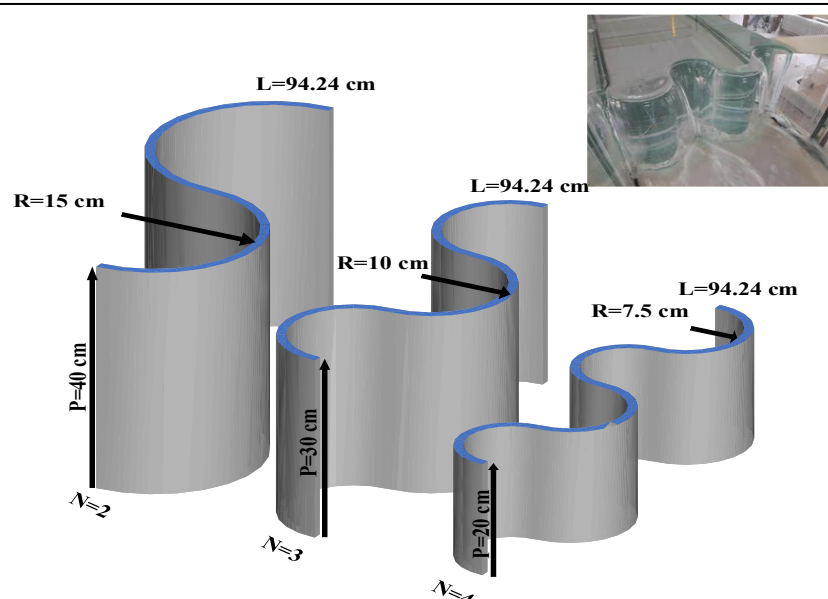
$$Cd = F_1(N, R, L, B, H_T, P, g, \rho, \mu, \delta) \quad (2)$$

در این رابطه، N معرف تعداد سیکل کنگره، R نشانگر شعاع انحنا، B بیانگر پهنای کانال، و P نماد ارتفاع سرریز است. همچنین، ρ ، μ و σ به ترتیب به چگالی سیال، ضریب لزجت دینامیکی و نیروی کشش سطحی دلالت دارند. به دلیل غیرقابل تراکم بودن جریان در سرریزها، اثر تغییرات جرم مخصوص (ρ) قابل صرف نظر است. افزون بر این، تأثیرات ناشی از لزجت و کشش سطحی، در وضعیتی که جریان ماهیت آشفته به خود گرفته و هد هیدرولیکی مستقر بر تاج سرریز، مطابق با رهنمود رو و شکالا^۲ (۱۹۷۱)، از سه سانتی‌متر فزونی یابد، می‌توان از آن‌ها صرف نظر نمود. بر این اساس، و با استفاده از تحلیل ابعادی، معادله (۲) به صورت زیر قابل بازنویسی خواهد بود:

$$Cd = F_2\left(N, \frac{R}{P}, \frac{L}{P}, \frac{B}{P}, \frac{H_T}{P}\right) \quad (3)$$

¹ Extreme Learning Machine

² Rao & Shukla



شکل ۲. پارامترهای موثر بر ضریب دبی سرریز کنگره‌ای نیم‌دایره طرح هارمونی

در پژوهش حاضر و در راستای توسعه مدل‌های محاسبات نرم، مجموعه داده‌های آزمایشگاهی به منظور برآورد ضریب تخلیه (Cd) در سرریزهای کنگره‌ای با پلان دایره‌ای مستخرج از پژوهش‌های ارهام نمازی و مظفری^۱ (۲۰۲۳) و یل‌دیز و همکاران^۲ (۲۰۲۴) استفاده شده است. این مجموعه داده‌ها، شامل بر ۲۱۵ آزمایش انجام شده در پژوهش تحت سه ارتفاع متفاوت برای سرریز (P) به ترتیب معادل ۲۰، ۳۰ و ۴۰ سانتی‌متر) و با سه تعداد مختلف از سیکل (N برابر با ۲، ۳ و ۴)، و همچنین داده‌های تکمیلی حاصل از مطالعات نمازی و مظفری^۳ (۲۰۲۳)، برای ارتفاع ثابت سرریز $P=15$ سانتی‌متر که به صورت تک سیکل ($N=1$) به تعداد ۱۸ آزمایش مبنای تحلیل‌های این پژوهش را تشکیل داده‌اند. بدین ترتیب، مجموعه‌ای بالغ بر ۲۳۳ نتیجه آزمایشگاهی، به منظور توسعه و اعتبارسنجی مدل‌های مبتنی بر محاسبات نرم، گردآوری شد. از این مجموعه، ۱۷۵ نمونه (معادل ۷۵ درصد از کل داده‌ها) به فرایند آموزش مدل‌ها اختصاص یافته و ۵۸ نمونه باقیمانده (معادل ۲۵ درصد) جهت آزمون و ارزیابی کارایی مدل‌های توسعه یافته، مورد استفاده قرار گرفته است. در مرحله بعد، بهینه‌سازی بیزی با استفاده از پنج زیرمجموعه (fold) از داده‌های آموزش اجرا شده و برای هر ترکیب از پارامترها، مقدار RMSE محاسبه می‌شود. چنانچه مقدار تابع هدف (RMSE) به حد مطلوب برسد، ترکیب پارامترها به عنوان ساختار نهایی مدل انتخاب می‌گردد. در نهایت، مدل بهینه شده با داده‌های آزمون ارزیابی شده و نتایج پیش‌بینی و عملکرد مدل با مدل‌های پایه مقایسه می‌شود. جدول (۱) شاخص‌های آماری تفکیک شده برای هر دو مجموعه داده (آموزش و آزمون) را نمایش می‌دهد. شکل (۳)، توزیع فراوانی (هیستوگرام) هر یک از متغیرهای مستقل و همچنین، ماتریس همبستگی میان این متغیرها و ضریب تخلیه (Cd) را ارائه می‌دهد. به طور کلی، این نمودار نشان می‌دهد همبستگی خطی ضعیفی بین متغیرها و Cd وجود دارد، بنابراین پیش‌بینی دقیق بر اساس روش‌های کلاسیک رگرسیون دشوار است.

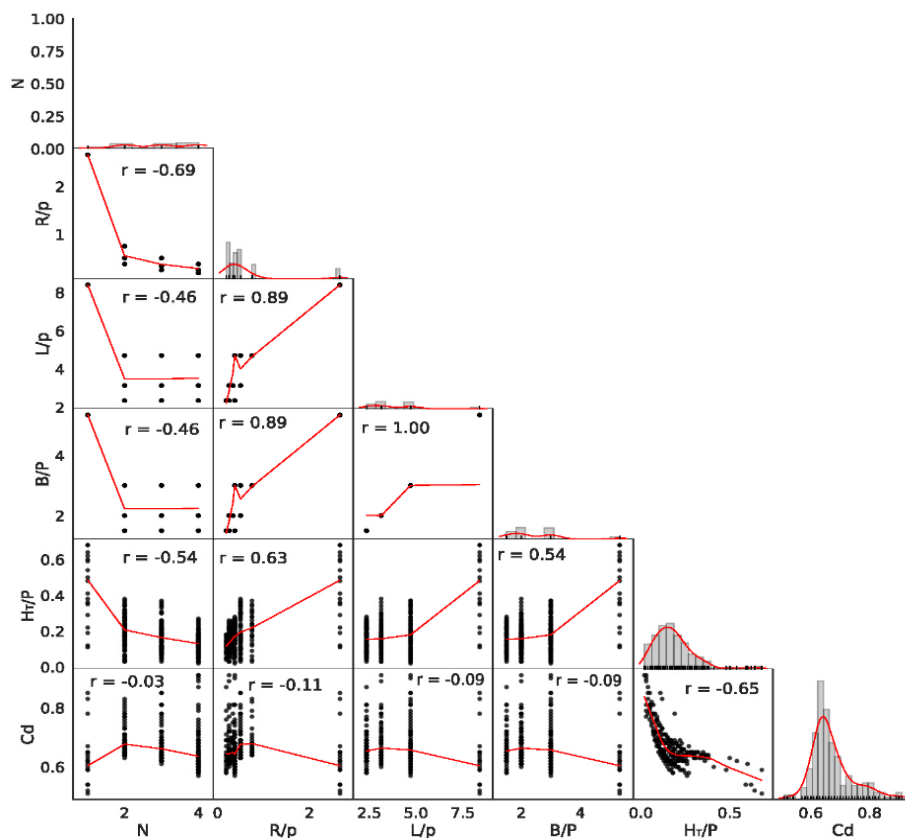
^۱ Arham Namazi & Mozaffari

^۲ Yıldız et al

^۳ Arham Namazi & Mozaffari

جدول ۱. آماره‌های توصیفی داده‌های تخصیص داده‌شده در مرحله آموزش و اعتبارسنجی مدل‌های یادگیری ماشین

مرحله	شاخص	cd	N	R/p	L/p	B/p	H/p
آموزش	کمینه	۰/۵۱۰	۱	۰/۱۸۷	۲/۳۵۶	۱/۵۰۰	۰/۰۲۰
	بیشینه	۰/۹۱۰	۴	۲/۶۶۷	۸/۴۰۰	۵/۳۳۳	۰/۶۸۰
	میانگین	۰/۶۷۲	۲/۹۰۳	۰/۵۴۶	۳/۷۹۴	۲/۴۱۴	۰/۱۸۷
	انحراف معیار	۰/۰۷۱	۰/۹۵۱	۰/۵۹۶	۱/۵۴۸	۰/۹۸۳	۰/۱۲۰
اعتبارسنجی	کمینه	۰/۵۴۰	۱	۰/۱۸۷	۲/۳۵۶	۱/۵	۰/۰۳۰
	بیشینه	۰/۸۹۰	۴	۲/۶۶۷	۸/۴۰۰	۵/۳۳۳	۰/۶۲۰
	میانگین	۰/۶۷۲	۲/۷۵۹	۰/۶۶۵	۴/۱۷۳	۲/۶۵۵	۰/۱۹۲
	انحراف معیار	۰/۰۶۹	۰/۹۷۹	۰/۷۰۸	۱/۷۳۱	۱/۰۹۸	۰/۱۱۳

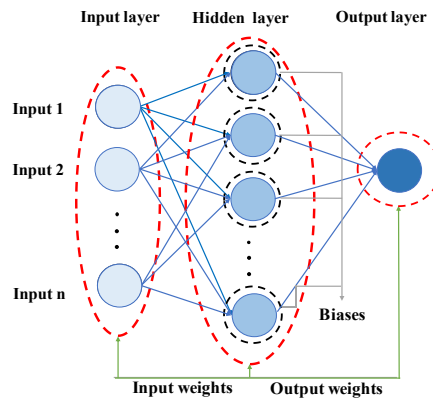


شکل ۳. نمودار هیستوگرام و ماتریس همبستگی بین همه متغیرها

۲. ماشین یادگیری افراطی (ELM)

شکل کلی ماشین یادگیری افراطی (ELM) در شکل (۴) نمایش داده شده است. این مدل شامل سه لایه ورودی، پنهان و خروجی است. همچنین، تعداد نورون‌های لایه ورودی با تعداد ورودی‌های مسئله مطابقت دارد و تعداد نورون‌های لایه خروجی نیز با تعداد خروجی‌های مسئله همخوانی دارد که در مسئله مدل‌سازی سرریز کنگره ای مطالعه حاضر ضریب دبی می باشد. همان‌طور که در شکل (۴) با یک خروجی واحد نشان داده شده است. لایه پنهان و لایه ورودی از طریق وزن‌های ورودی به یکدیگر متصل شده‌اند. علاوه بر این، خروجی لایه پنهان نیز با استفاده از وزن‌ها و بایاس‌های ورودی آن محاسبه می‌شود. این فرایند منجر به

تشکیل ماتریسی به نام «ماتریس خروجی لایه پنهان» می گردد که از مقادیر تصادفی این دو ماتریس و تابع فعال سازی مناسب حاصل می شود.



شکل ۴. معماری روش ELM

در هنگام آموزش مدل $\{(x_i, y_i)\}_{i=1}^N$ برای مسئله ای با یک خروجی و N نمونه، خروجی های متناظر به صورت $y_i \in \mathbb{R}$ و ورودی ها $x_i \in \mathbb{R}^n$ در نظر گرفته می شوند. با فرض اینکه شبکه عصبی چندلایه مخفی (HLNN) دارای L باشد و $g(x)$ به عنوان تابع فعال سازی (AF) به کار رود، شکل ریاضی روش ماشین یادگیری افراطی (ELM) برای برقراری رابطه میان پارامترهای ورودی و خروجی به شرح زیر ارائه می شود (عظیمی و شیرینی، ۲۰۲۱):

$$\sum_{i=1}^L \beta_i g(W_i j + b_i) = y_j, \quad j=1, 2, \dots, N \quad (4)$$

در این معادله، متغیرهای ورودی با نماد x_i نشان داده شده اند و خروجی متناظر با این ورودی ها با y_i نمایش داده می شود. همچنین، بایاس لایه پنهان در شاخص i با پارامتر b_i مشخص می گردد. لازم به ذکر است که تمامی بایاس های لایه های پنهان به صورت تصادفی مقداردهی اولیه می شوند.

در فرایند بهینه سازی مدل یادگیری افراطی ELM، دو ابرپارامتر اصلی شامل تعداد نرون های لایه پنهان^۲ و نوع تابع فعال سازی^۳ به منظور بهبود دقت پیش بینی مورد ارزیابی و تنظیم قرار گرفتند. پارامتر n_hidden بیانگر تعداد نرون ها در لایه میانی مدل است که نقش مهمی در ظرفیت یادگیری و پیچیدگی مدل ایفا می کند، درحالی که تابع فعال سازی نوع نگاشت غیرخطی بین لایه ها را مشخص می سازد و بر توان مدل در استخراج ویژگی های پنهان داده ها اثرگذار است. برای پارامتر تعداد نرون های لایه پنهان، بازه ای بین ۱۰ تا ۵۰۰ انتخاب شد که طیف گسترده ای از ظرفیت مدل را شامل می شود. در مورد تابع فعال سازی نیز سه تابع متداول و پرکاربرد شامل توابع سیگموئید (sigm)، تانژانت هیپربولیک (tanh) و خطی (lin) مدنظر قرار گرفتند.

۳. الگوریتم تقویت گرادیان شدید (XGBoost)

الگوریتم XGBoost نوعی روش یادگیری جمعی^۴ است که با ترکیب چندین درخت تصمیم ضعیف، یک مدل قدرتمند و دقیق می سازد. مطابق شکل (۵)، داده های ورودی (x_i, y_i) به مدل داده می شوند تا اولین درخت تصمیم $f_1(x)$ ساخته شود (چن و گسترین^۵، ۲۰۱۶؛ و همکاران^۶، ۲۰۲۲). خروجی این درخت معمولاً پیش بینی اولیه ای از مقادیر هدف است. در ادامه، خطای

¹ Azimi and Shiri

² n_hidden

³ activation function

⁴ Ensemble

⁵ Chen and Guestrin

⁶ Wu et al

باقی مانده^۱ بین مقدار واقعی y_i و پیش‌بینی شده محاسبه شده و در مرحله بعد به‌عنوان داده ورودی برای ساخت درخت بعدی $f_2(x)$ استفاده می‌شود. این روند به‌صورت تکراری ادامه می‌یابد و هر درخت جدید سعی می‌کند خطای باقی مانده از مدل قبلی را کاهش دهد. بنابراین مدل نهایی به شکل مجموع وزن دار تمام درخت‌های یادگیرنده ضعیف خواهد بود:

$$f_i^{(t)} = \sum_{k=1}^t f_k(x_i) = f_i^{(t-1)} + f_t(x_i) \quad (5)$$

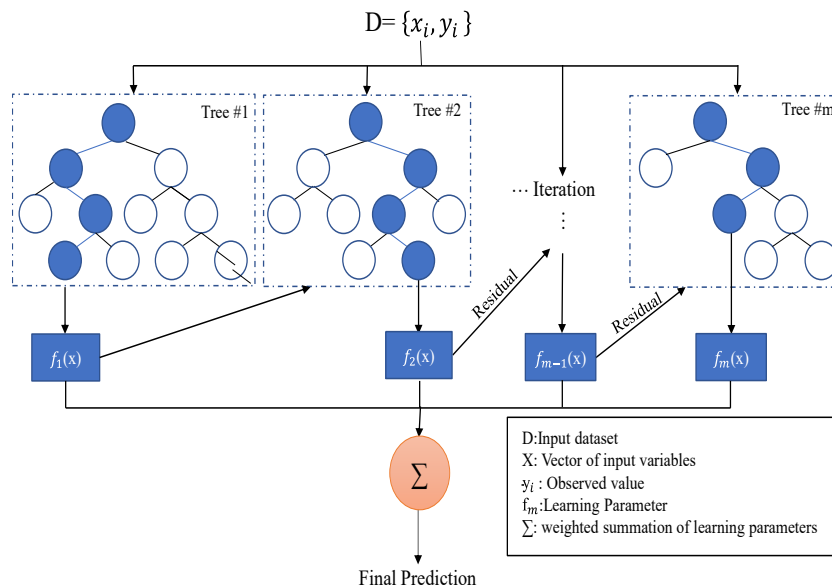
در این رابطه، $f_t(x_i)$ یادگیرنده جدید در مرحله t است، $f_i^{(t-1)}$ پیش‌بینی مرحله قبلی و x_i ورودی مدل است. همان‌طور که در نمودار ۵ دیده می‌شود، در هر مرحله تکرار، مدل جدید $f_t(x)$ براساس باقیمانده‌ی مدل قبلی ساخته می‌شود تا به بهبود پیش‌بینی کمک کند. تابع هدف مدل به شکل زیر تعریف می‌شود:

$$Obj^{(t)} = \sum_{k=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^t \Omega(f_i) \quad (6)$$

در اینجا l تابع هزینه است، n تعداد داده‌ها و Ω یک جمله منظم‌سازی برای جلوگیری از بیش‌برازش و کنترل پیچیدگی مدل است. جمله منظم‌سازی به‌صورت زیر است:

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \|\omega\|^2 \quad (7)$$

که در آن ω بردار امتیاز گره‌های برگ، λ ضریب منظم‌سازی و γ حداقل بهبود موردنیاز برای تقسیم بیشتر گره‌ها است. در نهایت، خروجی مدل به‌صورت ترکیب وزن دار تمامی درخت‌ها به‌دست آمده و به‌عنوان پیش‌بینی نهایی نمایش داده می‌شود.



شکل ۵. فلوچارت مدل *XGBoost* (پیرایی و همکاران^۲، ۲۰۲۳)

^۱ Residual

^۲ Piraei et al

در فرایند بهینه‌سازی مدل XGBoost با استفاده از الگوریتم بهینه‌سازی بیزی^۱، شش ابرپارامتر کلیدی شامل عمق درخت‌ها^۲، نرخ یادگیری^۳، تعداد درخت‌ها^۴، درصد نمونه‌برداری^۵، نسبت انتخاب ویژگی‌ها^۶ و پارامتر جریمه تقسیم^۷ مورد ارزیابی و تنظیم قرار گرفتند. این ابرپارامترها به ترتیب نقش تعیین‌کننده‌ای در کنترل پیچیدگی مدل، جلوگیری از بیش‌برازش، و بهبود دقت پیش‌بینی ایفا می‌کنند. برای تعیین دامنه جستجوی هر یک از این ابرپارامترها، ابتدا با استناد به مطالعات پیشین (پیتروویک و همکاران^۸، ۲۰۲۲؛ کوک و همکاران^۹، ۲۰۲۱)، محدوده‌ای اولیه و نسبتاً وسیع برای هر ابرپارامتر در نظر گرفته شد. این محدوده به‌گونه‌ای انتخاب گردید که احتمال وجود مقدار بهینه در آن بالا باشد. سپس با ثابت نگه‌داشتن سایر ابرپارامترها در مقادیر پیش‌فرض، مقادیر مختلفی از هر ابرپارامتر هدف در فواصل منظم مورد ارزیابی قرار گرفت.

در هر آزمون، مدل XGBoost بر مبنای مقدار انتخابی از ابرپارامتر موردنظر آموزش داده شد و ضریب تعیین (R^2) میان خروجی مدل و داده‌های آزمایشگاهی محاسبه گردید. چنانچه مقدار R^2 از ۰/۷۵ بیشتر می‌بود، مقدار مربوطه به‌عنوان مقدار قابل قبول ذخیره می‌شد. در نهایت، بیشینه و کمینه مقادیر ذخیره‌شده برای هر ابرپارامتر، به ترتیب به‌عنوان کران بالایی و پایینی دامنه جستجو در فرایند بهینه‌سازی تعیین شدند.

جدول ۲. محدوده پارامترهای قابل تنظیم مدل XGBoost

مؤلفه	محدوده تغییرات
عمق درخت‌ها	۳ تا ۲۰
نرخ یادگیری	۰.۰۱ تا ۰.۵
تعداد درخت‌ها	۵۰ تا ۱۰۰۰
درصد نمونه‌برداری	۰/۳ تا یک
نسبت انتخاب ویژگی‌ها	۰/۳ تا یک
پارامتر جریمه تقسیم	صفر تا ۱۰

این دامنه‌ها به‌عنوان ورودی در الگوریتم بهینه‌سازی بیزی تعریف شده و موجب تسریع در همگرایی مدل به سمت ترکیب بهینه از ابرپارامترها گردیدند. روند دستیابی به توسعه مدل بهینه‌سازی‌شده XGBoost از کتابخانه sklearn و bayes_opt در زبان برنامه‌نویسی پایتون حاصل شد.

۴. آنالیز حساسیت

۴-۱. روش موریس

روش موریس که در سال ۱۹۹۱ معرفی شد، با هدف تحلیل حساسیت و شناسایی پارامترهای کلیدی مدل طراحی شده است (موریس^{۱۰}، ۱۹۹۱). اساس این روش بر نمونه‌گیری تصادفی و اجرای مکرر شبیه‌سازی است، به طوری‌که در هر تکرار تنها یکی از پارامترها تغییر می‌کند. اثر اولیه تغییر هر پارامتر به‌صورت رابطه (۸) محاسبه می‌شود

¹ Bayesian Optimization

² Max_depth

³ learning_rate

⁴ N_estimators

⁵ Subsample

⁶ Colsample_bytree

⁷ Gamma

⁸ Petrovic et al

⁹ Koc et al

¹⁰ Morris

$$d_i(X) = \frac{f(x_1, \dots, x_{i-1}, x_i + \Delta, x_{i+1}, \dots, x_n) - f(x_1, \dots, x_{i-1}, x_i, \dots, x_n)}{\Delta} \quad (8)$$

در اینجا، مقدار Δ از میان مجموعه‌ای شامل مقادیر $\left\{\frac{1}{p-1}, \frac{2}{p-1}, \dots, 1 - \frac{1}{p-1}\right\}$ انتخاب می‌شود که در آن p بیانگر تعداد سطوح هر پارامتر است. بردار $(X_1, \dots, X_{i-1}, X_i, \dots, X_n)$ یک نمونه تصادفی از فضای پارامترها به‌شمار می‌رود. دو شاخص آماری برای ارزیابی اهمیت پارامترها معرفی شده‌اند: میانگین اثر اولیه μ و انحراف معیار σ . شاخص μ نمایانگر تأثیر کلی یک پارامتر است و اگر مقدار آن بالا باشد، پارامتر تأثیر زیادی دارد؛ و بالعکس. انحراف معیار بالا در σ نشان‌دهنده تعاملات پیچیده و رفتار غیرخطی در مدل است. برای بهبود تحلیل، نسخه‌های اصلاح شده از میانگین اثر اولیه نیز با عنوان μ^* به‌صورت رابطه (۹) پیشنهاد شده است (کینگ و پررا^۱، ۲۰۱۳).

$$\mu_i^* = \frac{1}{R} \left| \sum_{j=1}^R d_i(X^{(j)}) \right| \quad (9)$$

۴-۲. روش سوپول

روش سوپول یک تحلیل حساسیت سراسری^۲ است که بر پایه تجزیه واریانس خروجی مدل به اجزای نسبت‌داده‌شده به هر ورودی بنا شده است (سوپول^۳، ۱۹۹۳). واریانس کل خروجی مدل به‌صورت رابطه (۱۰) نمایش داده می‌شود:

$$V(y) = \sum_{i=1}^n V_i + \sum_{i \leq j \leq n} V_{ij} + \dots + \sum_{i \leq l} V_{l \dots n} \quad (10)$$

در این رابطه، $V(y)$ نشانگر مجموع واریانس کل خروجی مدل است. همچنین n تعداد کل پارامترهای مدل را نشان می‌دهد. شاخص‌های تحلیل حساسیت در مرتبه اول، دوم و کل، از طریق روابط (۱۱) تا (۱۳) قابل تعیین هستند.

$$S_i = \frac{V_i}{V(y)} \quad (11)$$

$$S_{ij} = \frac{V_{ij}}{V(y)} \quad (12)$$

$$S_{Ti} = \frac{V_{\sim i}}{V(y)} \quad (13)$$

در این معادلات، $V_{\sim i}$ نمایانگر سهم تمامی پارامترها به‌استثنای پارامتر i در واریانس خروجی مدل است. برای محاسبه این شاخصه ۱ از روش مونت کارلو استفاده می‌شود. شاخص S_i بالا نشان‌دهنده تأثیر مستقیم زیاد پارامتر، و S_{Ti} بالا نشان‌دهنده وابستگی گسترده خروجی به آن پارامتر (مستقیم یا غیرمستقیم) است.

۴-۳. روش آنتروپی

این روش، از آنتروپی و اطلاعات متقابل برای سنجش میزان وابستگی بین ورودی‌ها و خروجی مدل استفاده می‌شود (لی و همکاران^۴، ۲۰۱۳؛ زنگ و همکاران^۵، ۲۰۱۲). آنتروپی هر متغیر x به‌صورت رابطه (۱۴) تعریف می‌شود:

$$H(x) = -\sum_j P_j \ln P_j \quad (14)$$

¹ King and Perera

² Global sensitivity analysis

³ Sobol

⁴ Li et al

⁵ Zeng et al

اطلاعات متقابل بین دو متغیر x و y نیز به صورت رابطه (۱۵) و شاخص نرمال شده وابستگی به صورت رابطه (۱۶) محاسبه می شود:

$$I(x,y)=H(x)+H(y)-H(x,y) \quad (15)$$

$$U(x,y)=2 \times \frac{I(x,y)}{H(x)+H(y)} \quad (16)$$

در شاخص U ، مقدار صفر نشان دهنده استقلال کامل و مقدار ۱ بیانگر وابستگی کامل است. این روش امکان تحلیل حساسیت را حتی در شرایط وجود عدم قطعیت و رفتارهای غیرخطی فراهم می کند (میشرا و همکاران^۱، ۲۰۰۳).

۴-۴. روش SHAP

روش SHAP یک رویکرد مستقل از مدل برای تفسیر پیش بینی های حاصل از هر مدل یادگیری ماشین است (لوندبرگ و لی^۲، ۲۰۱۷). این روش با بهره گیری از نظریه بازی ها و تبیین های محلی، سهم هر ویژگی را در پیش بینی نهایی بر اساس اثرات نهایی (حاشیه ای) آن توزیع می کند. به منظور اطمینان از توزیع منصفانه مشارکت ویژگی ها، مقادیر SHAP با استفاده از رابطه (۱۷) محاسبه می شوند.

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(n-|S|-1)!}{n!} [v_x(S \cup \{i\}) - v_x(S)] \quad (17)$$

در اینجا، $\phi_i(v)$ نشان دهنده سهم یا مشارکت ویژگی i است. N مجموعه ای از تمام ویژگی های موجود در مدل بوده و n بیانگر تعداد کل ویژگی ها است. S زیرمجموعه ای از ویژگی ها بدون در نظر گرفتن i است، درحالی که $v_x(S)$ پیش بینی مورد انتظار برای زیرمجموعه S را نشان می دهد. عبارت $[v_x(S \cup \{i\}) - v_x(S)]$ تغییر در مقدار پیش بینی به واسطه حضور ویژگی i را نشان می دهد.

۵. معیارهای ارزیابی

به منظور ارزیابی مدل های مورد استفاده در این تحقیق از شش شاخص های آماری ضریب تعیین (R^2)، ریشه میانگین مربعات خطا (RMSE)، میانگین درصد خطای مطلق (MAPE)، خطای نسبی مطلق (RAE)، ضریب تعیین نش (NS) و شاخص توافق (IA) به شرح زیر استفاده گردید:

$$R^2 = \frac{[\sum_{i=1}^N (O_i - \bar{O}_i)(P_i - \bar{P}_i)]^2}{\sum_{i=1}^N (O_i - \bar{O}_i)^2 \sum_{i=1}^N (P_i - \bar{P}_i)^2} \quad (16)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (O_i - P_i)^2}{N}} \quad (17)$$

$$MAPE = \frac{100}{N} \sum_{i=1}^N \left| \frac{P_i - O_i}{O_i} \right| \quad (18)$$

¹ Mishra et al

² Lundberg and Lee

$$RAE = \frac{\sum_{i=1}^N |P_i - O_i|}{\sum_{i=1}^N |O_i - \bar{O}_i|} \quad (۱۹)$$

$$NS = 1 - \frac{\sum_{i=1}^N (P_i - O_i)^2}{\sum_{i=1}^N (O_i - \bar{O}_i)^2} \quad (۲۰)$$

$$IA = 1 - \frac{\sum_{i=1}^N (P_i - O_i)^2}{\sum_{i=1}^N (|P_i - \bar{O}_i| + |O_i - \bar{O}_i|)^2} \quad (۲۱)$$

در رابطه‌های فوق O_i مقادیر ضریب دبی مشاهداتی (واقعی)، P_i مقادیر ضریب دبی پیش‌بینی شده، $\bar{O}_i$ میانگین مقادیر ضریب دبی مشاهداتی، $\bar{P}$ میانگین مقادیر ضریب دبی پیش‌بینی شده و N تعداد مشاهدات هستند. لازم به ذکر است مقادیر R^2 نزدیک یک و مقادیر RMSE، RAE و MAPE نزدیک به صفر، نشان‌دهنده دقت بیشتر و عملکرد بهتر مدل هستند.

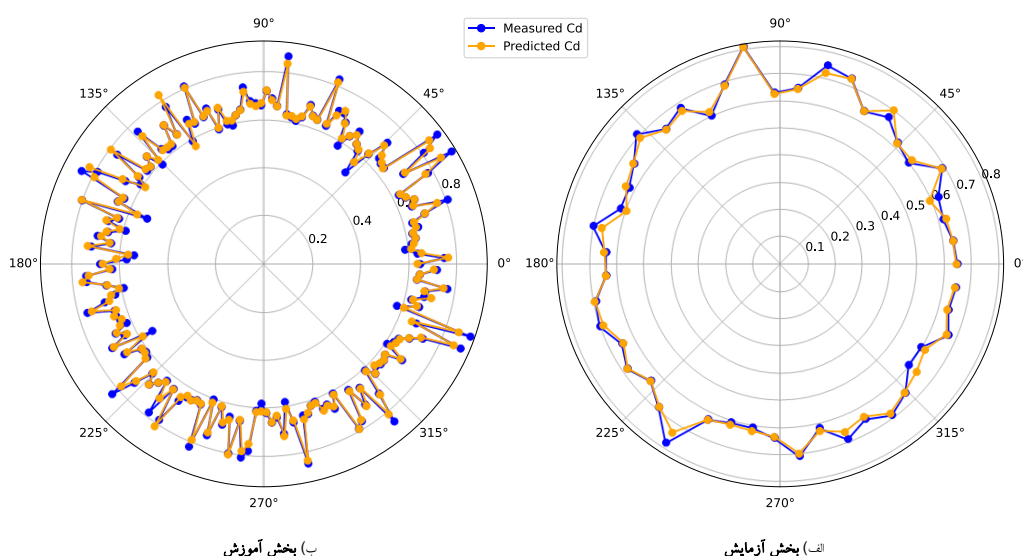
یافته‌های پژوهش

۱. نتایج مدل Xgboost

به منظور مدل‌سازی ضریب دبی (Cd)، همان‌گونه که بیان گردید، الگوریتم XGBoost با استفاده از رویکرد ترکیبی بهینه‌سازی بیزی و اعتبارسنجی متقابل^۱ پنج‌تایی بهینه‌سازی شد. در فرایند بهینه‌سازی، شش پارامتر کلیدی مدل شامل عمق درخت‌ها، نرخ یادگیری، تعداد درخت‌ها، درصد نمونه‌برداری، نسبت انتخاب ویژگی‌ها و پارامتر جریمه تقسیم تنظیم گردیدند. مقدار بهینه‌یافته این پارامترها در جدول (۳) آورده شده است. مطابق شکل (۶) ارزیابی عملکرد مدل نهایی نشان می‌دهد که الگوریتم XGBoost قادر است دقت بالایی در پیش‌بینی ضریب دبی در این نوع از سرریزها ارائه دهد. براین اساس، مقدار ضریب مقدار ضریب تعیین (R^2) برای مجموعه آموزش معادل ۰/۹۹۱۱ و برای مجموعه آزمون برابر با ۰/۹۲۳ به دست آمد. همچنین، مقدار ریشه میانگین مربعات خطا (RMSE) به ترتیب ۰/۰۰۷ برای داده‌های آموزش و ۰/۰۱۲۷ برای داده‌های آزمون گزارش شد. نتایج نشان‌دهنده انطباق مناسب بین مقادیر واقعی و پیش‌بینی‌شده‌ی Cd می‌باشند.

در نمودارهای قطبی شکل (۶)، ترسیم‌شده برای داده‌های آموزش و آزمون، روند هم‌راستای نقاط پیش‌بینی‌شده با مقادیر واقعی کاملاً مشهود است که گواهی بر دقت مدل در شناسایی الگوهای حاکم بر داده‌ها دارد. در مجموع، این مدل می‌تواند به عنوان ابزاری مؤثر و قابل اعتماد در برآورد ضریب دبی سرریزهای نیمه‌دایره‌ای مورد استفاده قرار گیرد.

^۱ K Fold



شکل ۶. نمودارهای قطبی مربوط به مقادیر پیش‌بینی شده مدل *Xgboost* در برابر مقادیر واقعی ضریب دبی (*Cd*) برای سریزهای نیم‌دایره الف) بخش آزمایش و ب) بخش آموزش

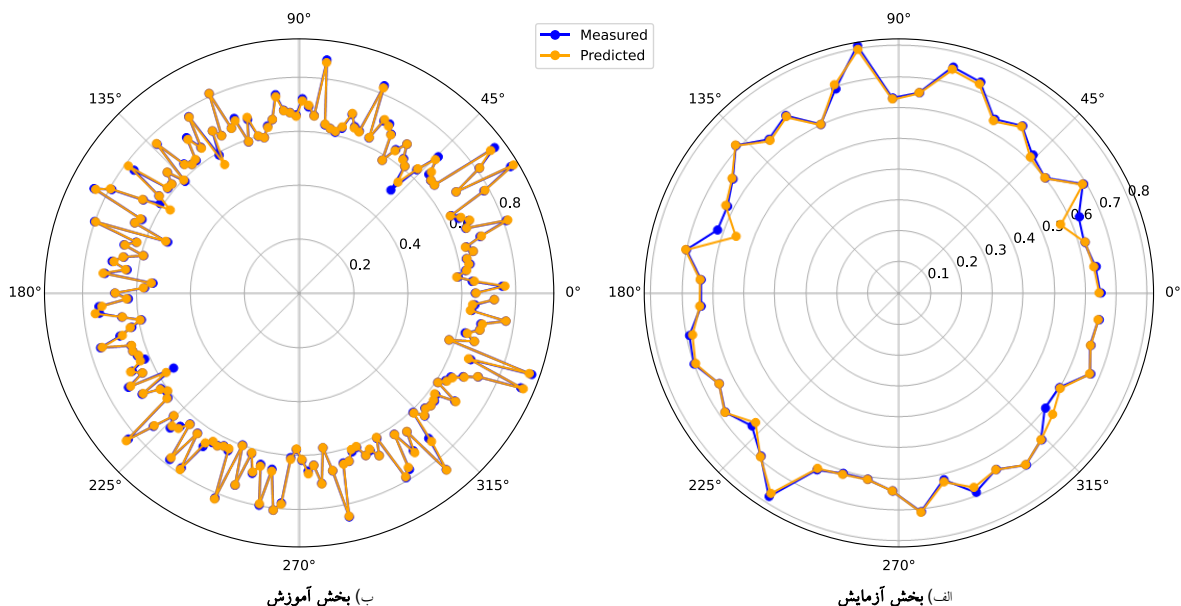
۲. نتایج مدل *ELM*

به منظور مدل‌سازی ضریب دبی (*Cd*)، مدل یادگیری افراطی با استفاده از رویکرد بهینه‌سازی بیزی و اعتبارسنجی متقابل پنج‌تایی بهینه‌سازی گردید. در این فرایند، دو پارامتر کلیدی شامل تعداد نرون‌های لایه پنهان و نوع تابع فعال‌سازی شبکه مورد ارزیابی قرار گرفت که در جدول (۳) مقادیر آنها نشان داده شده است. در شکل (۷) ارزیابی عملکرد مدل نهایی نشان داد که ساختار *ELM* با ترکیب پارامترهای بهینه‌شده، دقت قابل‌توجهی در پیش‌بینی مقادیر *Cd* دارد. به طور خاص، مقدار ضریب تعیین (R^2) برای مجموعه آموزش معادل ۰/۹۱ و برای مجموعه آزمون برابر با ۰/۸۷۹ محاسبه گردید. همچنین، مقادیر ریشه میانگین مربعات خطا (*RMSE*) برای داده‌های آموزش و آزمون به ترتیب برابر با ۰/۰۰۷ و ۰/۰۱۲۷ بوده است.

جدول ۳. پارامترهای بهینه شده مدل‌های یادگیری ماشین

مدل	پارامترهای بهینه شده
<i>XGBOOST</i>	۳ = عمق درخت‌ها
	۰/۱۶ = نرخ یادگیری
	۱۹۵ = تعداد درخت‌ها
	۸۸ = درصد نمونه‌برداری
	۰/۹۴ = نسبت انتخاب ویژگی‌ها
	۰/۰۰۵۷ = پارامتر جریمه تقسیم
<i>ELM</i>	$Sigm$ = تابع فعال‌ساز
	۲۲ = تعداد نرون

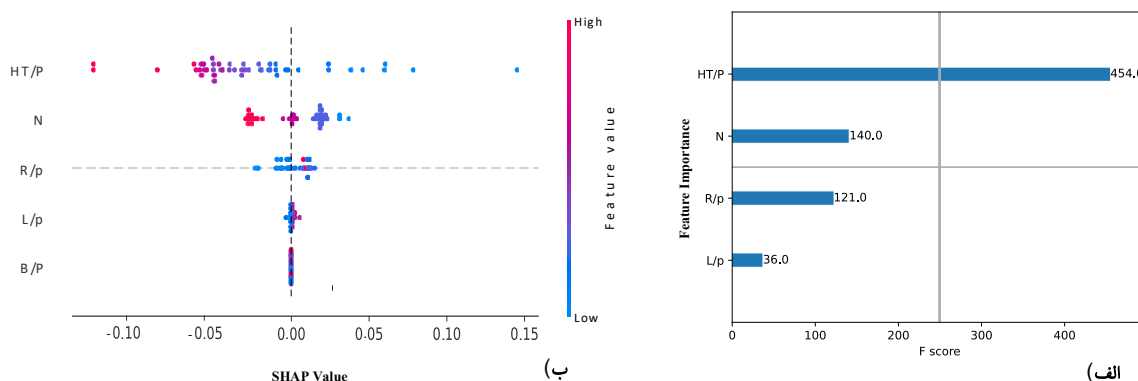
مقایسه نمودارهای قطبی در شکل (۷) حاصل از مقادیر واقعی و پیش‌بینی شده برای دو بخش آموزش و آزمون نیز بیانگر هم‌راستایی و انطباق بالای مدل با داده‌های واقعی است. الگوهای ترسیمی در این نمودارها نشان می‌دهند که مدل ELM با ساختار بهینه همانند مدل Xgboost توانسته است روند تغییرات Cd را با دقت بالا بازسازی کند.



شکل ۷. نمودارهای قطبی مربوط به مقادیر پیش‌بینی شده مدل ELM در برابر مقادیر واقعی ضریب دبی (Cd) برای سریزهای نیم‌دایره الف) بخش آزمایش و ب) بخش آموزش

۳. نتایج تحلیل حساسیت

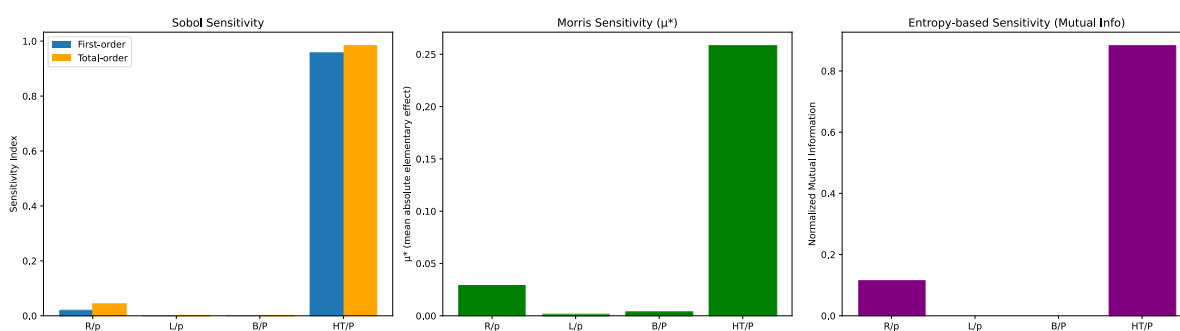
شکل (۸-الف) ارائه شده تحلیل دقیقی از اهمیت و تأثیر ویژگی‌ها در مدل XGBoost ارائه می‌دهند. بر اساس این نمودار که محور افقی تعداد دفعاتی که ویژگی در تقسیم‌گرهای درخت استفاده شده است، اهمیت متغیر H_T/P با امتیاز ۴۵۴ بالاترین نقش را در پیش‌بینی مدل ایفا کرده و مهم‌ترین عامل در تصمیم‌گیری مدل بوده است. پس از آن، متغیرهای N با امتیاز ۱۴۰، R/P با امتیاز ۱۲۱ و در نهایت L/P با امتیاز ۳۶ قرار دارند. این مقادیر نشان می‌دهند که مدل به ترتیب بیشترین توجه را به H_T/P ، سپس N و R/P و کمترین توجه را به L/P داشته است. در شکل ۸ ب، نمودار SHAP نیز برای بررسی دقیق‌تر و تفسیرپذیری مدل به کار رفته است. در این نمودار، مقدار SHAP برای هر ویژگی میزان تأثیر آن را بر خروجی مدل نشان می‌دهد. رنگ هر نقطه بر اساس مقدار واقعی ویژگی در داده‌ها تعیین شده است؛ رنگ آبی نشان‌دهنده مقادیر پایین و رنگ قرمز نشان‌دهنده مقادیر بالای ضریب دبی (Cd) است. همانند شکل (۸-ب)، ویژگی H_T/P بیشترین گستردگی در مقدار SHAP را دارد که بیانگر تأثیرگذاری بالای آن بر پیش‌بینی مدل است؛ به‌طوری که مقادیر پایین این ویژگی (آبی‌رنگ) عمدتاً تأثیر مثبت و مقادیر بالا (قرمز) تأثیر منفی بر خروجی مدل دارند. سایر ویژگی‌ها مانند N و R/P نیز اثراتی قابل توجه دارند؛ اما در مقایسه با H_T/P کمتر هستند. ویژگی B/P تقریباً تأثیر خاصی در خروجی مدل نداشته و مقدار SHAP آن حول صفر متمرکز است.



شکل ۸. آنالیز حساسیت با الف) روش درختی Xgboost و ب) روش SHAP

در ادامه برای شناسایی متغیرهای اثرگذار بر ضریب دبی (Cd)، از سه روش معتبر تحلیل حساسیت شامل سوبول (تحلیل واریانس جهانی)، روش موریس (اثر اولیه میانگین) و روش مبتنی بر آنترپوی (اطلاع متقابل) استفاده شد. نتایج هر سه روش نیز، برتری متغیر HT/P را به عنوان مهم ترین پارامتر ورودی به وضوح نشان می دهند. مطابق شکل (۹) در روش سوبول، شاخص حساسیت مرتبه اول برای HT/P برابر با حدود ۰/۹۹ و شاخص کل نیز بسیار نزدیک به همان مقدار بوده است. این موضوع بیانگر آن است که تقریباً تمام واریانس پاسخ مدل را می توان تنها با تغییر در این پارامتر توجیه کرد و تأثیر متقابل آن با سایر پارامترها بسیار ناچیز است. در مقابل، پارامترهای R/p، L/p و B/p تأثیر بسیار ناچیزی دارند به طوری که شاخص کل آن ها به سختی از مقدار ۰/۰۲ تجاوز می کند. در روش Morris نیز، مقدار میانگین اثر مطلق اولیه (μ^*) برای HT/P برابر با حدود ۰/۲۶ گزارش شده که اختلاف چشمگیری با سایر پارامترها دارد. R/p با مقدار حدودی ۰/۰۳ دومین پارامتر مؤثر محسوب می شود، اما تأثیر آن در مقایسه با HT/P بسیار پایین تر است. در نهایت، در روش اطلاع متقابل نیز بیشترین حساسیت به HT/P اختصاص دارد به طوری که مقدار نرمال شده آن حدود ۰/۸۷ بوده و R/p با فاصله زیاد و مقدار حدودی ۰/۱۳ در رتبه دوم قرار می گیرد، درحالی که L/p و B/p تقریباً اثر قابل ذکری ندارند.

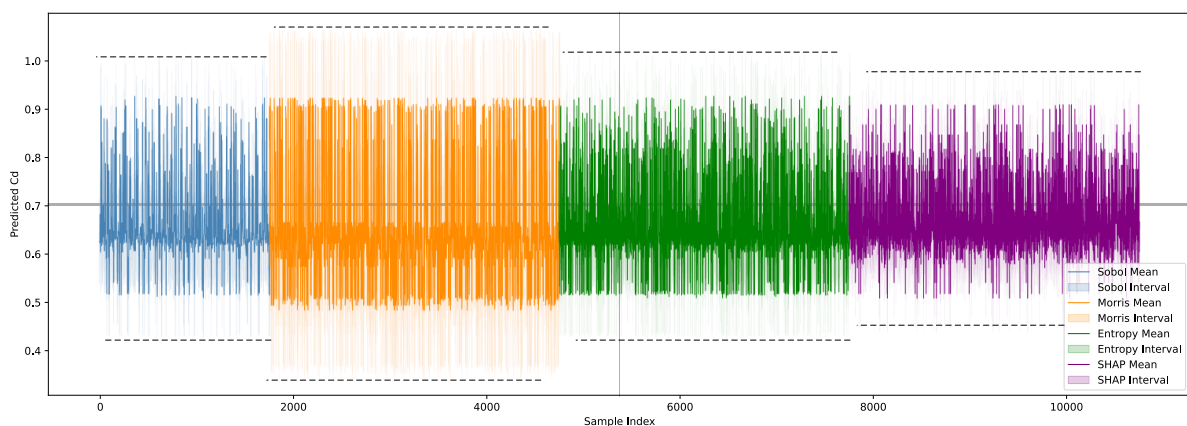
مقایسه کلی این پنج روش نشان می دهد که به طور هم راستا پارامتر HT/P را به عنوان مؤثرترین ورودی شناسایی کرده اند که نشان از پایایی نتایج تحلیل حساسیت و اعتبار مدل مورد استفاده دارد. همچنین، اتفاق نظر این روش ها می تواند مبنای تصمیم گیری برای کاهش بعد مسئله و طراحی آزمایش های بعدی قرار گیرد.



شکل ۹. آنالیز حساسیت با استفاده از روش های آنترپوی، موریس و سوبول

در شکل (۱۰) نمودار چهار روش مختلف نمونه گیری برای تحلیل حساسیت مدل XGBoost را مقایسه می کند. محور افقی معرف ایندکس نمونه برداری است که برای هر چهار روش آنالیز حساسیت برابر ۳۰۰۰ در نظر گرفته شده است. محور عمودی مقدار پیش بینی شده برای متغیر هدف (Cd) است. برای هر روش، میانگین پیش بینی ها به صورت خط پررنگ و نوار بازه

(interval) که نمایانگر تغییرات ($\pm$ انحراف معیار) است با رنگی روشن تر ترسیم شده است. به طور کلی، توزیع خروجی مدل در همه روش ها تقریباً متمرکز در بازه $0/4$ تا یک قرار دارد، اما شدت نوسانات و پراکندگی در هر روش آنالیز حساسیت متفاوت است. در مقایسه بین روش ها، مشاهده می شود که روش سوبول (آبی) و آنتروپی (سبز) دارای بازه های نوسان متعادل هستند که نشان از پایداری نسبی پیش بینی ها در این نمونه گیری ها دارد. روش موریس (نارنجی) دارای گستره نوسان بزرگ تری است که ممکن است ناشی از ماهیت تصادفی تر یا ساختار وابسته تری به پارامترهای ورودی باشد. در مقابل، آنالیز SHAP (بنفش) اگرچه مبتنی بر خروجی مدل است، اما بازه های نوسان آن محدودتر به نظر می رسد که نشان دهنده این است که ویژگی های انتخاب شده توسط SHAP در محدوده ای پایدار اثرگذار هستند. این نتیجه می تواند حاکی از این باشد که SHAP نه تنها ابزاری برای تفسیر مدل است؛ بلکه برای ارزیابی حساسیت نیز رفتار مدل را به شکلی هدفمند منعکس می کند. در نهایت، استفاده ترکیبی از این روش ها می تواند تحلیل جامع تری از پایداری و حساسیت مدل نسبت به ورودی ها فراهم کند.



شکل ۱۰. بررسی عملکرد چهار روش آنالیز حساسیت در پایداری و پیش بینی ضریب دبی

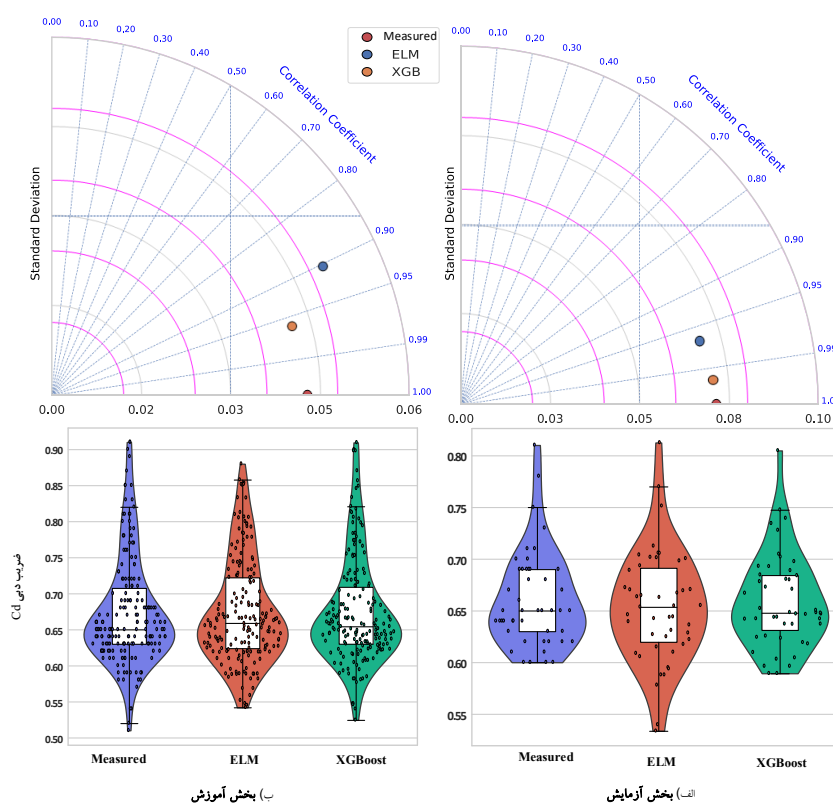
بحث

در این مطالعه، عملکرد دو مدل یادگیری ماشین XGBoost و ELM در پیش بینی مقادیر ضریب دبی (Cd) تحت نمونه گیری های حاصل از چهار روش آنالیز حساسیت سوبول، موریس، آنتروپی و SHAP مورد بررسی قرار گرفت. هر چهار روش مهم ترین پارامتر تأثیرگذار را متغیر H_T/P معرفی نمودند. برای ارزیابی عملکرد مدل های یادگیری ماشین، از نمودار تیلور و ویولن مطابق شکل (۱۱) استفاده شد. با تحلیل دقیق نمودارهای تیلور برای ارزیابی عملکرد مدل های پیش بینی ضریب دبی (Cd) در دو بخش آموزش و آزمایش، می توان تصویری جامع از دقت، همبستگی و پایداری هر مدل به دست آورد. در نمودار تیلور مربوط به بخش آموزش (نمودار سمت چپ)، داده های اندازه گیری شده (Measured) با انحراف معیار حدود $0/48$ و ضریب همبستگی یک به عنوان مرجع قرار گرفته اند. مدل XGBoost با قرارگیری بسیار نزدیک به نقطه مرجع و داشتن ضریب همبستگی بالا در حدود $0/97$ و انحراف معیار تقریباً برابر با داده های واقعی، عملکرد بسیار دقیقی در یادگیری از داده ها نشان داده است. در مقابل، مدل ELM که با نقطه آبی مشخص شده، ضریب همبستگی پایین تری در حدود $0/93$ دارد و موقعیت آن دورتر از داده های واقعی است که بیانگر دقت کمتر در بازتولید الگوهای آموزشی است.

در نمودار سمت راست که مربوط به بخش آزمایش است، مجدداً مدل XGBoost عملکرد بهتری از خود نشان داده است. نقطه نارنجی آن دارای ضریب همبستگی بسیار بالا (بیش از $0/95$) و انحراف معیار بسیار نزدیک به مقدار واقعی است. در حالی که

مدل ELM با انحراف معیار بالاتر و ضریب همبستگی پایین تر (تقریباً ۰/۹ یا کمتر)، فاصله بیشتری از نقطه مرجع دارد. این نشان می دهد که XGBoost نه تنها در مرحله آموزش دقت بالایی دارد و رفتار ضریب دبی را دقیق تر پیش بینی می کند.

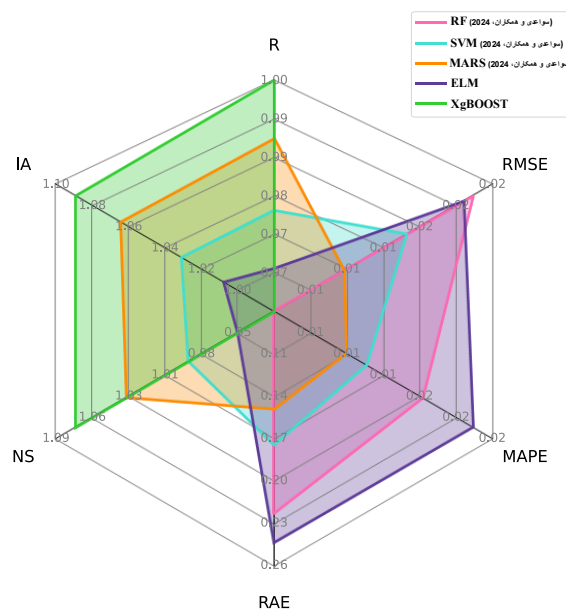
نمودارهای ویولن نمایش داده شده توزیع مقادیر ضریب دبی (Cd) را برای داده های اندازه گیری شده و دو مدل یادگیری ماشین ELM و XGBoost در دو بخش آزمایش (شکل ۱۱-الف) و آموزش (شکل ۱۱-ب) به تصویر می کشند. در بخش آموزش، مقدار میانه ی Cd برای داده های واقعی حدود ۰/۷ بوده و توزیع نسبتاً متراکمی در این محدوده دیده می شود. مدل ELM در این بخش میانه ای حدود ۰/۶۹ دارد و توزیع آن گرایش بیشتری به ناحیه پایین تر (حدود ۰/۶۵ تا ۰/۷) دارد، در حالی که مدل XGBoost با میانه ای تقریباً برابر با ۰/۷ و توزیعی متراکم در اطراف آن، بسیار نزدیک به داده های واقعی عمل کرده است. همچنین، دامنه ی تغییرات در XGBoost به نسبت متقارن تر و نرمال تر است که نشان دهنده دقت و ثبات بهتر این مدل در مرحله آموزش است. در بخش آزمایش، توزیع داده های واقعی در محدوده ۰/۶۵ تا ۰/۷ متمرکز است و میانه ی آن حدود ۰/۶۷ است. مدل ELM در این بخش با میانه ای در حدود ۰/۶۶ و دامنه ی نسبتاً پهن و کشیده ای که به سمت مقادیر بالاتر و پایین تر انحراف دارد، دقت کمتری نسبت به بخش آموزش نشان می دهد. از سوی دیگر، مدل XGBoost در بخش آزمایش نیز با میانه ای نزدیک به ۰/۶۷ و توزیعی متمرکزتر، تطابق بهتری با داده های واقعی دارد. پهنای کمتر نمودار ویولن برای XGBoost به ویژه در مرکز داده ها (۰/۶۵ تا ۰/۷) نشان دهنده همگرایی بهتر پیش بینی ها در اطراف مقدار واقعی است. به طور کلی، تحلیل این دو نمودار نشان می دهد که مدل XGBoost هم در مرحله آموزش و هم در مرحله آزمون، در بازتولید و پیش بینی ضریب دبی عملکرد بهتری نسبت به مدل ELM داشته و با پراکندگی کمتر، میانه نزدیک تر به داده های واقعی و توزیع متمرکزتر، گزینه ای دقیق تر و پایدارتر برای مدل سازی به شمار می رود.



شکل ۱۱. دیاگرام تیپور و ویولن عملکرد مدل ELM و XGBoost در (الف) بخش آزمایش و (ب) بخش آموزش

همچنین مقایسه روش‌های مختلف آنالیز حساسیت نیز نشان می‌دهد که روش‌های SHAP و Entropy در مدل XGBoost، بازه‌های پیش‌بینی باریک‌تری تولید کرده‌اند که نشان‌دهنده اعتماد بالای مدل به ورودی‌های حاصل از این روش‌هاست. در مقابل، روش مورس بازه‌های نوسانی‌تر و گسترده‌تری داشتند که ممکن است به دلیل ماهیت تصادفی‌تر یا وابستگی بیشتر به برهم‌کنش‌های متغیرها در این روش باشد. به طور خاص، SHAP به‌عنوان دقیق‌ترین روش از نظر محدود کردن عدم قطعیت شناخته شد و به نظر می‌رسد که هماهنگی بالایی با مدل XGBoost دارد. بر اساس نمودار رادار ارائه‌شده در شکل (۱۲)، عملکرد سه مدل هوش مصنوعی شامل جنگل تصادفی^۱، ماشین بردار پشتیبان^۲ و مدل رگرسیون غیرپارامتری چندمتغیره اسپیلاین^۳ حاصل از نتایج مطالعات سواعدی و همکاران^۴ (۲۰۲۴) با نتایج حاصل از مدل‌های ELM و XgBoost مورد مقایسه قرار گرفت. نتایج با استفاده از شاخص‌های (R، MAPE، RMSE، IA، NS و RAE) ارزیابی شده‌اند. همان‌گونه که مشاهده می‌شود در بین مدل‌ها، مدل XgBOOST عملکرد بسیار بهتری در بیشتر شاخص‌ها از خود نشان داده است. به طور خاص، این مدل دارای مقادیر نزدیک به یک در شاخص‌های NS، IA و R بوده که نشان از دقت بسیار بالا در بازتولید داده‌های واقعی دارد.

در نتیجه، می‌توان نتیجه‌گیری کرد که مدل XgBOOST با اختلاف قابل‌توجهی بهترین عملکرد را در بین مدل‌های بررسی‌شده داشته و گزینه مناسبی برای پیش‌بینی ضریب دبی جریان در این مطالعه است. مدل‌های RF و ELM ضعیف‌ترین عملکرد را نشان داده‌اند، درحالی‌که مدل‌های MARS و SVM در جایگاه میانی قرار دارند. این تحلیل می‌تواند به‌عنوان مرجع تصمیم‌گیری در انتخاب مدل‌های پیش‌بینی در مطالعات آینده مورد استفاده قرار گیرد.



شکل ۱۲. بررسی عملکرد مدل‌های ELM و Xgboost با سایر مدل‌های هوش مصنوعی انجام شده در پیش‌بینی ضریب دبی سرریز نیم‌دایره

¹ Random forest

² Support Vector Machine

³ multivariate adaptive regression spline

⁴ Savaedi et al

نتیجه گیری

پیش بینی دقیق ضریب دبی در سرریزهای نیم دایره طرح هارمونیک، برای طراحی و عملکرد بهینه سیستم های هیدرولیکی از اهمیت بالایی برخوردار است. به دلیل ارتباط پیچیده بین ضریب دبی و پارامترهای مؤثر بر آن، معادلات تجربی در پیش بینی دقیق این ضریب با محدودیت هایی مواجه هستند. در مطالعه حاضر، دو مدل ترکیبی جدید یادگیری ماشین با عنوان ELM و XGBoost توسعه داده شده است که برای مدل سازی ضریب دبی در این نوع از سازه ها به کار گرفته شد. این دو مدل جدید، مزایای الگوریتم تجمعی تقویتی و افراطی را با الگوریتم بهینه سازی بیزی ترکیب می کنند. به منظور کاهش خطای پیش بینی در مدل های یادگیری ماشین، ابتدا تحلیل اهمیت ویژگی ها با استفاده از روش چهار روش انجام شد تا میزان تأثیر نسبی هر پارامتر ورودی مشخص گردد.

در مجموع، ترکیب مدل XGBoost با روش های نمونه گیری SHAP می تواند بهترین گزینه برای پیش بینی دقیق و پایدار Cd باشد. این ترکیب نه تنها بازه های پیش بینی باریکی ارائه می دهد، بلکه نسبت به داده های نمونه برداری شده مقاومت بیشتری در برابر نویز و نوسانات دارد. این نتایج می تواند راهنمای مناسبی برای انتخاب مدل و روش نمونه گیری در مطالعات آبی مرتبط با تحلیل حساسیت و پیش بینی ویژگی های وابسته به عدم قطعیت باشد. همچنین نتایج نشان داد مدل XGBoost با وجود ساختار پیشرفته و قابلیت های تنظیم پارامتر، عملکرد نسبتاً ضعیف تری نسبت به ELM داشته است. در نمودار تیلور، فاصله مدل XGB از نقطه مرجع در هر دو بخش بیشتر از ELM است که این موضوع بر دقت کمتر آن در پیش بینی دلالت دارد.

ملاحظات اخلاقی

پیروی از اصول اخلاق پژوهش

نویسندگان اصول اخلاقی را در انجام و انتشار این پژوهش علمی رعایت نموده اند و این موضوع مورد تأیید همه آنهاست.

تعارض منافع

بنا بر اظهار نویسندگان این مقاله تعارض منافع ندارد.

سپاسگزاری

در پایان، از حمایت های مالی شورای پژوهشی دانشگاه شهید چمران اهواز کمال تشکر را داریم (GN: SCU.WHI403.43525).

References

- Arham Namazi, F. S., & Mozaffari, J. (2023). Investigation of labyrinth weirs discharge coefficient with the same length. *Flow Measurement and Instrumentation*, 94, 102468. <https://doi.org/10.1016/j.flowmeasinst.2023.102468>
- Azimi, H., & Shiri, H. (2021). Assessment of ice-seabed interaction process in clay using extreme learning machine. *International Journal of Offshore and Polar Engineering*, 31(04), 411–420. <https://doi.org/10.17736/ijope.2021.mt31>
- Bilhan, O., Emiroglu, M. E., Miller, C. J., & Ulas, M. (2018). The evaluation of the effect of nappe breakers on the discharge capacity of trapezoidal labyrinth weirs by ELM and SVR approaches. *Flow Measurement and Instrumentation*, 64, 71–82. <https://doi.org/10.1016/j.flowmeasinst.2018.10.009>
- Bonakdari, H., Gholami, A., Mosavi, A., Kazemian-Kale-Kale, A., Ebtehaj, I., & Azimi, A. H. (2020). A novel comprehensive evaluation method for estimating the bank profile shape and dimensions of stable channels using the maximum entropy principle. *Entropy*, 22(11), 1218. <https://doi.org/10.3390/e22111218>
- Borghei, S. M., & Parvaneh, A. (2011). Discharge characteristics of a modified oblique side weir in subcritical flow. *Flow Measurement and Instrumentation*, 22(5), 370–376. <https://doi.org/10.1016/j.flowmeasinst.2011.04.009>
- Borisova, N. M., & Belikov, V. V. (2023). Optimization of Spillway Design of Reconstructed Shavan Dam Using Numerical 2D-Modeling. *Power Technology and Engineering*, 57(1), 67–74. Scopus. <https://doi.org/10.1007/s10749-023-01624-7>
- Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. <https://doi.org/10.1145/2939672.2939785>
- Dutta, D., Mandal, A., & Afzal, M. S. (2020). Discharge performance of plan view of multi-cycle W-form and circular arc labyrinth weir using machine learning. *Flow Measurement and Instrumentation*, 73, 101740. <https://doi.org/10.1016/j.flowmeasinst.2020.101740>
- Emami, S., Parsa, J., Emami, H., & Abbaspour, A. (2021). Investigation of Discharge Coefficient of Triangular Duckbill Labyrinth Weirs Using Fluent Software and Gray Wolf and Elections Algorithms. *Civil Infrastructure Researches*, 6(2), 107–121. <https://doi.org/10.22091/cer.2021.6942.1250>
- Ghaderi, D., Ebrahimnezhadian, H., & Mollazadeh, M. (2024). Three-dimensional analysis of the performance of circular stepped spillways in the skimming flow regime. *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, 46(7). <https://doi.org/10.1007/s40430-024-05004-8>
- Haghiabi, A. H., Parsaie, A., & Ememgholizadeh, S. (2018). Prediction of discharge coefficient of triangular labyrinth weirs using Adaptive Neuro Fuzzy Inference System. *Alexandria Engineering Journal*, 57(3), 1773–1782. <https://doi.org/10.1016/j.aej.2017.05.005>
- Hussein, B. S., & Jalil, S. A. (2024). Influence of Labyrinth Side Weir Shape Modification on the Hydrodynamic Performance: Experimental and Numerical Study. *Arabian Journal for Science and Engineering*, 50, 12881–12902. <https://doi.org/10.1007/s13369-024-09563-9>
- Ikinciogullari, E. (2024). Energy dissipation performance of labyrinth and harmonic stepped spillways. *Journal of Hydroinformatics*, 26(10), 2668–2682. <https://doi.org/10.2166/hydro.2024.221>
- Koc, K., Ekmekcioğlu, Ö., & Gurgun, A. P. (2021). Integrating feature engineering, genetic algorithm and tree-based machine learning methods to predict the post-accident disability

- status of construction workers. *Automation in Construction*, 131, 103896. <https://doi.org/10.1016/j.autcon.2021.103896>
- Kisi, O., Emin Emiroglu, M., Bilhan, O., & Guven, A. (2012). Prediction of lateral outflow over triangular labyrinth side weirs under subcritical conditions using soft computing approaches. *Expert Systems with Applications*, 39(3), 3454–3460. <https://doi.org/10.1016/j.eswa.2011.09.035>
- Li, J., Duan, Q. Y., Gong, W., Ye, A., Dai, Y., Miao, C., Di, Z., Tong, C., & Sun, Y. (2013). Assessing parameter importance of the Common Land Model based on qualitative and quantitative sensitivity analysis. *Hydrology and Earth System Sciences*, 17(8), 3279–3293.
- Li, S., Shen, G., Parsaie, A., Li, G., & Cao, D. (2024). Discharge modeling and characteristic analysis of semi-circular side weir based on the soft computing method. *Journal of Hydroinformatics*, 26(1), 175–188. <https://doi.org/10.5194/hessd-10-2243-2013>
- Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30. <https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html>
- Mishra, S., & Knowlton, R. G. (2003). Testing for input-output dependence in performance assessment models. *Proceedings of the 10th International High-Level Radioactive Waste Management Conference (IHLRW)*, March 30–April 2, 2003, Las Vegas, Nevada, 882–887. <https://doi.org/10.1023/a:1007058325258>
- Morris, M. D. (1991). Factorial Sampling Plans for Preliminary Computational Experiments. *Technometrics*, 33(2), 161–174. <https://doi.org/10.1080/00401706.1991.10484804>
- King, D. M., & Perera, B. J. C. (2013). Morris method of sensitivity analysis applied to assess the importance of input variables on urban water supply yield—A case study. *Journal of hydrology*, 477, 17–32. <https://www.sciencedirect.com/science/article/pii/S0022169412008918>
- Petrovic, A., Strumberger, I., Antonijevic, M., Jovanovic, D., Mladenovic, D., & Chabbra, A. (2022). Firefly-xgboost approach for pedestrian detection. *2022 IEEE Zooming Innovation in Consumer Technologies Conference (ZINC)*, 197–202. <https://ieeexplore.ieee.org/abstract/document/9840700/>
- Piraei, R., Afzali, S. H., & Niazkar, M. (2023). Assessment of XGBoost to Estimate Total Sediment Loads in Rivers. *Water Resources Management*, 37(13), 5289–5306. <https://doi.org/10.1007/s11269-023-03606-w>
- Rao, S. S., & Shukla, M. K. (1971). Characteristics of flow over weirs of finite crest width. *Journal of the Hydraulics Division*, 97(11), 1807–1816. <https://doi.org/10.1061/JYCEAJ.0003138>
- Safarrazavi Zadeh, M., Esmaeili Varaki, M., & Biabani, R. (2021). Experimental study on flow over sinusoidal and semicircular labyrinth weirs. *ISH Journal of Hydraulic Engineering*, 27(sup1), 304–313. <https://doi.org/10.1080/09715010.2019.1644679>
- Sangsefidi, Y., & Ghodsian, M. (2018). Investigation of effects of entrance channel walls on the hydraulic performance of arced labyrinth weirs. <https://www.sid.ir/paper/256636/en>
- Savaedi, A., Zayeri, M., Ghomeshi, M., & Daryaei, M. (2024b). Application of Artificial Intelligence in Predicting the Discharge Coefficient of Labyrinth Weirs with Harmonic Plans. *Irrigation and Drainage Structures Engineering Research*, 25(96), 82–61. <https://doi.org/10.22092/idser.2025.368037.1602>
- Shafiei, S., Najarchi, M., & Shabanlou, S. (2020). Simulation of labyrinth weir discharge coefficient by modern artificial intelligence models. *Modares Civil Engineering Journal*, 20(1), 204–218. https://mcej.modares.ac.ir/browse.php?a_id=35509&sid=16&slc_lang=en

- Simsek, O., Gumus, V., & Ozluk, A. (2023). Prediction of discharge coefficient of the trapezoidal broad-crested weir flow using soft computing techniques. *Neural Computing and Applications*, 35(24), 17485–17499. <https://doi.org/10.1007/s00521-023-08615-9>
- Sobol, I. M. (1993). Sensitivity estimates for nonlinear mathematical models. *Math. Model. Comput. Exp.*, 1(4), 407–414. <https://www.scirp.org/reference/referencespapers?referenceid=2692473>
- Suleimany, J. M. F. S., Aurahman, T. H., & Mamand, B. S. (2022). Flow simulation over semicircular labyrinth weir using ANSYS-fluent. *Tikrit Journal of Engineering Sciences*, 29(1), 83–98. <https://doi.org/10.25130/tjes.29.1.7>
- Wu, J., Ma, D., & Wang, W. (2022). Leakage Identification in Water Distribution Networks Based on XGBoost Algorithm. *Journal of Water Resources Planning and Management*, 148(3), 04021107. [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0001523](https://doi.org/10.1061/(ASCE)WR.1943-5452.0001523)
- Yıldız, A., Marti, A. İ., & Göğüş, M. (2024). The hydraulic investigation of harmonic plan weirs. *Flow Measurement and Instrumentation*, 95, 102512. <https://doi.org/10.1016/j.flowmeasinst.2023.102512>
- Zare, H., Vaghefi, M., Mahmoudi, A., & Behroozi, A. M. (2023). Experimental Exploration of Flow Hydraulics and Discharge Coefficient for an Inclined Circular Labyrinth Weir. *Water Resources Management*, 37(11), 4521–4536. <https://doi.org/10.1007/s11269-023-03577-y>
- Zeng, X., Wang, D., & Wu, J. (2012). Sensitivity analysis of the probability distribution of groundwater level series based on information entropy. *Stochastic Environmental Research and Risk Assessment*, 26, 345–356. <https://doi.org/10.1007/s00477-012-0556-2>