

## Research Paper

# Performance of Support Vector Machines, Random Forest, and GMDH Methods for Predicting Flow Rate from Rectangular Flap Gates

Bahareh Behdarvandi<sup>1</sup>,Mohammad Reza Zayeri<sup>2\*</sup>,Mehdi Ghomeshi<sup>3</sup>, Mehdi Daryaei<sup>4</sup>

<sup>1</sup>Master's student in water civil engineering and hydraulic structures, Faculty of Water and Environmental Engineering, Shahid Chamran University of Ahvaz, Ahvaz, Iran. ([bahare.bd7@gmail.com](mailto:bahare.bd7@gmail.com))

<sup>2\*</sup>Assistant professor, Faculty of Water and Environmental Engineering, Shahid Chamran University of Ahvaz, Ahvaz, Iran. ([m.zayri@scu.ac.ir](mailto:m.zayri@scu.ac.ir))

<sup>3</sup>professor, Faculty of Water and Environmental Engineering, Shahid Chamran University of Ahvaz, Ahvaz, Iran. ([Ghomeshi@scu.ac.ir](mailto:Ghomeshi@scu.ac.ir))

<sup>4</sup>Associate professor, Faculty of Water and Environmental Engineering, Shahid Chamran University of Ahvaz, Ahvaz, Iran. ([m.daryaei@scu.ac.ir](mailto:m.daryaei@scu.ac.ir))



10.22125/iwe.2024.460044.1811

Received:

**June 24, 2024**

Accepted:

**November 9, 2024**

Available online:

**December 26, 2024****Keywords:****Flap Gates, automatic control discharge, irrigation networks, soft Computing, flow measurement****Abstract**

Installation and establishment of appropriate flow measurement structures, their proper operation, and collection of water consumption data play a crucial role in empowering network managers' decision-making, ensuring fair water distribution and transmission, and ultimately achieving water conservation. Flap or hanging gates are simple and inexpensive devices for both automatic control and flow measurement. In this research, the laboratory results obtained from a rectangular hinged gate structure, which is placed in a channel according to hydraulic and geometric specifications and leads to flow rate measurement, were used to develop machine learning models. To estimate the flow rate in this type of channel, models including Group Method of Data Handling (GMDH), Support Vector Machines (SVM), and Random Forest (RF) were employed. To this end, parameters such as water depth, channel width, and gate width, length, thickness, and weight were considered as input variables, and flow rate as the output (response) variable for modeling. The results showed that the Root Mean Square Error (RMSE) values for GMDH-, SVM-, and RF-based models were 0.024, 0.011, and 0.041, respectively, and the Coefficient of Determination (R<sup>2</sup>) values were 0.981, 0.996, and 0.955, respectively. A comparison between past research and the present results indicated the superiority of the SVM-based model over the other developed models. Water depth to channel width was identified as the most significant input data for the developed models

\* **Corresponding Author:** Mohammad Reza Zayeri

**Address:** Assistant professor of Water and Environmental Engineering faculty, Shahid Chamran University of Ahvaz, Ahvaz, Iran.

**Email:** [m.zayri@scu.ac.ir](mailto:m.zayri@scu.ac.ir)

**Tel:** 09364930766

## 1) Introduction

Accurate flow rate measurement is essential for various applications in hydraulic engineering, including water resource management, irrigation systems, and hydropower generation. Conventional methods for flow rate measurement, such as weirs and flumes, often require complex installations and may not be suitable for all situations. Flap gates, on the other hand, offer a simple and cost-effective solution for both flow control and measurement. Machine learning has emerged as a powerful tool for modeling complex relationships in various fields, including hydraulic engineering. Machine learning models can learn from experimental data and make predictions without explicit knowledge of the underlying physical relationships. In this study, we investigate the application of machine learning models for predicting the flow rate in rectangular flap gates channels.

## 2) Materials and Methods

To evaluate the accuracy of the proposed equations for estimating the discharge coefficient of hanging gates under free flow conditions, the experimental dataset of Gorgin and Vatankhah (2022), obtained in the Hydraulic Laboratory of the Department of Irrigation Engineering at the University of Tehran, was utilized for modeling purposes. The experimental results employed in this study were acquired in a flume with a constant width of 0.25 m, a height of 0.5 m, and a length of 12 m for two channel bed slopes of  $S = 0.005$  and horizontal ( $S = 0$ ). Since the discharge passing through the hanging gate is related to the geometrical characteristics of the channel and gate, as well as the hydraulic characteristics of the upstream flow depth, seven dimensionless input parameters derived from dimensional analysis in accordance with laboratory experiments were considered in this study for developing the GMDH, SVM, and RF models.

## 3) Results

Three machine learning models, namely Group Method of Data Handling (GMDH), Support Vector Machines (SVM), and Random Forest (RF), were developed to predict the flow rate in rectangular flap gates channels. To develop soft computing-based models, the collected data must first be divided into two categories: training and validation. In this study, 80% of the data (550 samples) was allocated to the training phase, while the remaining 20% (137 samples) was allocated to the validation of the developed models. Since the collected data does not have a time series nature, the allocation of data to each of the training and validation groups was done randomly. The results of the Random Forest model in the training and validation phases are presented in Figures (4) and (5), respectively. The statistical indices of the developed Random Forest model in the training phase are  $R^2 = 0.967$  and  $RMSE = 0.035$ , while in the validation phase, they are  $R^2 = 0.955$  and  $RMSE = 0.041$ . Sensitivity analysis was performed to investigate the impact of input parameters on the performance of the developed machine learning models. The results of the sensitivity analysis revealed that the relative water depth and gate opening angle

variables had the most significant influence on the estimation of flow rate, while the  $\frac{M}{\rho b^3}$  and  $\frac{t}{B}$  parameters had the least impact.

## 4) Discussion and Conclusion

This study investigated the application of soft computing models, including Support Vector Machines (SVM), Group Method of Data Handling (GMDH), and Random Forest (RF), for flow rate prediction in rectangular flap gates channels. The results demonstrated that all three developed soft computing models outperformed the non-linear regression model proposed by Gorgin and Vatankhah (2022) in terms of

accuracy in predicting flow rate in channels with hanging gates. The sensitivity analysis of the Random Forest model revealed that the relative water depth to channel width ratio ( $H/B$ ) was the most significant parameter influencing flow rate prediction, accounting for 72% of the total contribution. The least

influential parameter was the gate thickness ratio ( $\frac{t}{B}$ ), with a contribution of only 0.1%. Among the developed machine learning models, the SVM model demonstrated the most accurate flow rate prediction with  $R^2$  and RMSE values of 0.995 and 0.041, respectively, in the validation phase.

### Conflict of Interest

Authors declared no conflict of interest.

### 5\_Six important references

1. Gorgin, F., Vatankhah, A.R., 2022. Rectangular top-hinged plate as portable flow measuring device. *Water Supply* 22, 8637–8658.
2. Mahmoudi, B., Farhoudi, J., 2019. Discharge estimation of flap gate installed in a circular channel under free flow condition. *Iran. Water Res. J.* 13, 127–136.
3. Momeni, E., Nazir, R., Jahed Armaghani, D., Maizir, H., 2014. Prediction of pile bearing capacity using a hybrid genetic algorithm-based ANN. *Measurement* 57, 122–131. <https://doi.org/10.1016/j.measurement.2014.08.007>
4. Taylor, K.E., 2001. Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res. Atmospheres* 106, 7183–7192. <https://doi.org/10.1029/2000JD900719>
5. Zayeri, M., 2023. Discharge Prediction in Flumes with Trapezoidal Contraction by Machine Learning Techniques. *Irrig. Drain. Struct. Eng. Res.* 24, 55–70. <https://doi.org/10.22092/idser.2023.363054.1549>
6. Zhang, Y., Chiew, F.H.S., Li, M., Post, D., 2018. Predicting Runoff Signatures Using Regression and Hydrological Modeling Approaches. *Water Resour. Res.* 54, 7859–7878. <https://doi.org/10.1029/2018WR023325>



## عملکرد روش های ماشین بردار پشتیبان، جنگل تصادفی و GMDH در پیش بینی دبی جریان از دریچه های فلپ مستطیل شکل

بهاره بهداروندی<sup>۱</sup>، محمدرضا زایری<sup>۲\*</sup>، مهدی قمشی<sup>۳</sup> و مهدی دریائی<sup>۴</sup>

تاریخ ارسال: ۱۴۰۳/۰۴/۰۴

تاریخ پذیرش: ۱۴۰۳/۰۸/۱۹

مقاله پژوهشی

### چکیده

نصب و ایجاد سازه های مناسب اندازه گیری جریان، بهره برداری صحیح از آنها و جمع آوری اطلاعات مربوط به مصرف آب، نقشی اساسی در قدرت تصمیم گیری برای مدیران شبکه، عدالت در توزیع و انتقال آب و در نهایت صرفه جویی در مصرف آب دارد. یکی از تجهیزات ساده و ارزان برای کنترل خودکار و همچنین اندازه گیری جریان، دریچه های فلپ یا آویخته است. در این پژوهش از نتایج آزمایشگاهی به دست آمده از سازه دریچه مستطیلی لولای شکل که با قرار دادن در کانال مطابق با مشخصات هیدرولیکی و هندسی منجر به اندازه گیری دبی جریان میشود، جهت توسعه مدل های یادگیری ماشین مورد استفاده قرار گرفت. به منظور برآورد دبی جریان در این نوع از کلنال ها از مدل های شامل دسته بندی گروهی داده ها (GMDH)، ماشین بردار پشتیبان (SVM) و جنگل تصادفی (RF) استفاده گردید. بدین جهت از پارامترهای نظیر عمق آب، عرض کلنال همچنین عرض، طول، ضخامت و وزن دریچه به عنوان متغیرهای ورودی و دبی جریان به عنوان متغیر خروجی (پاسخ) جهت مدلسازی در نظر گرفته شد. نتایج نشان داد که مقدار آماره ریشه میانگین مربعات خطا (RMSE) برای مدل های مبتنی بر SVM، GMDH، RF به ترتیب، ۰/۰۲۴، ۰/۰۱۱ و ۰/۰۴۱ و مقدار آماره ضریب تعیین ( $R^2$ ) به ترتیب، ۰/۹۸۱، ۰/۹۹۶ و ۰/۹۵۵ به دست آمد. مقایسه بین تحقیقات گذشته با روابط کلاسیک و نتایج حاضر حاکی از برتری عملکرد مدل مبتنی بر SVM نسبت به سایر مدل های توسعه یافته بود با این وجود مدل GMDH به دلیل ارائه رابطه صریح بین متغیرهای ورودی و خروجی می تواند در تحقیقات آینده مورد استفاده قرار گیرد. همچنین نتایج آنالیز حساسیت نشان داد، عمق آب به عرض کانال و زاویه بازشدگی دریچه به عنوان مهم ترین داده های ورودی به مدل های یادگیری ماشین می باشند.

**واژه های کلیدی:** دریچه آویزان، کنترل خودکار، شبکه های آبیاری، محاسبات نرم، اندازه گیری دبی

<sup>۱</sup> دانشجوی کارشناسی ارشد، گروه سازه های آبی، دانشکده مهندسی آب و محیط زیست، دانشگاه شهید چمران اهواز، اهواز، ایران

[bahare.bd7@gmail.com](mailto:bahare.bd7@gmail.com)

<sup>۲</sup> استادیار، گروه سازه های آبی، دانشکده مهندسی آب و محیط زیست، دانشگاه شهید چمران اهواز، اهواز، ایران [m.zayri@scu.ac.ir](mailto:m.zayri@scu.ac.ir)

(نویسنده مسئول)

<sup>۳</sup> استاد، گروه سازه های آبی، دانشکده مهندسی آب و محیط زیست، دانشگاه شهید چمران اهواز، اهواز، ایران [Ghomeshi@scu.ac.ir](mailto:Ghomeshi@scu.ac.ir)

<sup>۴</sup> دانشیار، گروه سازه های آبی، دانشکده مهندسی آب و محیط زیست، دانشگاه شهید چمران اهواز، اهواز، ایران [m.daryaee@scu.ac.ir](mailto:m.daryaee@scu.ac.ir)

## مقدمه

با توجه به بحران کمبود آب تخصیص بهینه آن برای آبیاری کشاورزی به صورت کارآمد یک مسئله بحرانی در مدیریت آب کشاورزی است (Valipour, 2013). یکی از پارامتر مهم در مدیریت شبکه‌های آبیاری اندازه‌گیری دقیق دبی جریان می باشد که تأثیر بسیار مهمی بر تخصیص آب آبیاری دارد و اطلاع از آن، منجر به تقویت صرفه‌جویی در مصرف آب در بین کشاورزان می شود. ساختارهای ساده اندازه‌گیری دبی جریان نظیر سرریزها و دریچه‌ها در سال‌های اخیر گسترش یافته و انواع مختلفی از آنها توسعه یافته است. دریچه‌های آویخته<sup>۱</sup> از یک صفحه فلزی نازک تشکیل شده‌اند که به یک بازو متصل است. این بازو به دو نقطه لنگرگاه در دو طرف کانال متصل می‌شود. دریچه می‌تواند حول لنگرگاه تحت تأثیر نیروهای وارده باز و بسته شود. با افزودن یا کم کردن وزن دریچه می‌توان میزان جریان را کنترل کرد. با افزایش سطح آب در پشت دریچه، نیروهای ناشی از فشار آب وارد بر دریچه افزایش یافته و باعث باز شدن هرچه بیشتر دریچه می‌شوند. هنگامی که سطح آب در پشت دریچه افزایش می‌یابد، فشار آب وارد بر دریچه نیز افزایش می‌یابد. این فشار باعث باز شدن دریچه می‌شود. با افزودن وزن به دریچه، نیروی وارد بر آن افزایش می‌یابد. این نیرو باعث بسته شدن دریچه می‌شود. از مزایای دریچه‌های آویزان می‌توان به هزینه پایین ساخت و نصب، قابلیت انتقال رسوب بالادست، قابلیت نصب دائمی و یا موقت، خودکار بودن دریچه و هزینه تعمیر و نگهداری کم آن اشاره نمود (Yang et al., 2017).

با توجه به قابلیت‌های فوق، مطالعات آزمایشگاهی گوناگونی در مورد هیدرولیک جریان از زیر دریچه‌های آویخته در نقاط مختلف دنیا صورت گرفته است. Burt et al. (2001) مدلی را برای کنترل خودکار سطح آب در بالادست دریچه‌های آویخته مستطیلی شکل توسعه دادند. آنان با بهره‌گیری از مدل فیزیکی این نوع از دریچه‌ها در مقیاس کوچک، توزیع فشار روی دریچه را به دست آوردند و توانستند نیروی وارده به دریچه را با استفاده از یک

معادله خطی که تابع درجه باز شدگی دریچه باشد، تعیین کنند. (Replogle and Wahlin (2003) مطالعه‌ای را بر روی دریچه آویخته با مقطع دایره‌ای که در پایین دست زهکش نصب شده بود انجام دادند. آنان به این نتیجه رسیدند با پایین بودن وزن دریچه و بالا بودن سرعت جریان، افت انرژی ایجاد شده توسط دریچه، به صفر میل می‌کند و زمانی که وزن دریچه بیشتر و سرعت جریان کمتر باشد، میزان افت انرژی بیشتر می‌شود. پژوهش‌های انجام شده حاکی از این است که افت انرژی در دریچه آویخته نسبت به دیگر سازه‌های هیدرولیکی، که کاربرد یکسانی با این سازه دارند، کمتر است. Litrico et al. (2005) برآورد ضریب دبی جریان از دریچه آویخته با مقطع مستطیلی شکل را مورد مطالعه قرار دادند. آنان با استفاده از روش پایستگی برآورد لنگرهای وارد شده به لولای دریچه، معادلات رگرسیونی را برای تخمین دبی جریان در حالت رژیم جریان آزاد پیشنهاد دادند. (Mahmoudi and Farhoudi (2019) با روشهای تحلیلی با محاسبه دبی سرریز و معادله مومنوم به منظور تخمین دبی جریان عبوری از دریچه آویخته پرداختند. بررسی‌های آنان نشان داد که متوسط خطای نسبی در استفاده از معادلات توسعه یافته برای دریچه آویخته دایره‌ای بین ۴/۴۱ تا ۵/۶۶ درصد و در دریچه آویخته مربعی بین ۵/۵۵ تا ۶/۸۴ درصد متغیر است. Moghtaderi and Valizadegan (2021) در مطالعه‌ای برآورد دبی جریان را با استفاده از فرم رابطه دبی-اشل از دریچه‌های آویزان که در انتهای کانال‌های دوزنقه‌ای با شیب جانبی دیواره ثابت، مورد بررسی قرار دادند. مقایسه نتایج آزمایشگاهی با رابطه پیشنهادی آنان بر اساس معیار درصد متوسط خطای نسبی ۰/۲۲ درصد بوده است. در مطالعه‌ای توسط Ghaderinia et al. (2019) به منظور برآورد میزان دبی در کانال‌های دایره‌ای روباز افقی با استفاده از دریچه‌های آویزان نیم‌دایره‌ای، با ارائه رابطه آبگذری سازه موردنظر که تابعی از مشخصات هیدرولیکی جریان و مشخصات هندسی سازه است، رابطه‌ای صریح برای تخمین عمق بحرانی در کانال‌های دایره‌ای توسط

از دریچه های کشویی مستطیل شکل با استفاده از شبکه عصبی پیچشی پرداختند. نتایج آنان نشان میدهد که مدل CNN توسعه یافته می تواند با دقت الگوی جریان اطراف این سازه های را با مقادیر ضریب تبیین ( $R^2$ ) بیش از ۹۰٪ مدل سازی کند و به طور قابل توجهی هزینه های محاسباتی را کاهش دهد. (Kartal et al., 2024). بررسی ظرفیت دبی عبوری از سرریزهای H شکل مستطیلی با استفاده از مدل های SVM پرداختند. نتایج آنان معادله دبی عبوری پیشنهادی با استفاده از پارامترهای نسبت بازشدگی دریچه به عمق آب بالادست ( $a/y_1$ )، نسبت هد آب روی سرریز به عرض بازشدگی آن ( $h/b$ )، نسبت عرض بازشدگی سرریز به عرض کانال ( $b/B$ )، عدد فرود بالادست ( $Fr_1$ ) و نسبت فاصله عمودی تاج سرریز تا بالای دریچه ( $d/a$ ) با میانگین درصد خطای مطلق (MAPE) حدود ۵/۸۴ درصد را نشان میدهد.

بررسی مروری بر پیشینه موضوع نشان می دهد که مطالعه هیدرولیک جریان عبوری از دریچه های آویخته به منظور اندازه گیری دبی جریان عمدتاً بر اساس تحقیقات آزمایشگاهی انجام گرفته است و مدل سازی مبتنی بر علم داده نظیر روش های یادگیری ماشین جهت تفسیرپذیری روابط پیشنهادی دبی عبوری جریان مورد توجه واقع نشده است. این مهم با وجود پارامترهای هندسه مختلف این سازه نیازمند مطالعه بیشتر بوده تا علاوه بر ارائه روابط دقیق غیرخطی، سهم پارامترهای مهم ورودی در فرمول بندی ارزیابی گردد، لذا در این تحقیق به توسعه مدل های مبتنی بر دسته بندی گروهی داده ها (GMDH)، مدل ماشین بردار پشتیبان (SVM) و الگوریتم جنگل تصادفی (RF) در زبان برنامه نویسی پایتون جهت برآورد دبی جریان در دریچه های فلپ مستطیل شکل پرداخته شد.

## مواد و روش ها

### کانال اندازه گیری دبی جریان با دریچه آویخته

به منظور ارزیابی دقت معادلات پیشنهادی در برآورد ضریب دبی برای دریچه های آویخته در شرایط جریان آزاد، از مجموعه داده های آزمایشگاهی Gorgin and

آنها ارائه شد. رابطه به دست آمده با استفاده از داده های آزمایشگاهی (۲۵۴ سری داده) واسنجی شد. نتایج آنان نشان می دهد متوسط خطای برآورد دبی توسط رابطه به دست آمده، ۳/۶ درصد در صورت صرف نظر کردن از اثرات لزجت و سه درصد در صورت لحاظ کردن اثرات لزجت بیان گردید. همچنین آنان با توجه به پایین بودن متوسط و بیشینه خطای روابط ارائه شده برای صفحه نیم دایره ای آویزان، این سازه را به عنوان یک سازه کارآمد در راستای اندازه گیری دبی در کانال نیم دایره ای پیشنهاد نمودند. در مطالعه ای Rabiee Moghadam et al. (2018) به بررسی هیدرولیکی و طراحی دریچه فلپ اصلاح شده برای کنترل خودکار سطح آب پرداختند. آزمایشات مختلف آنان بر روی دریچه فلپ نشان داد که استفاده از زاویه بیشتر از ۴۵ درجه و نیز کاهش فاصله مرکز ثقل دریچه تا صفحه وجه بالادست دریچه باعث کاهش نوسانات سطح آب در دبی های مختلف میشود.

توسعه اخیر علم داده توانایی آن در حل مسائل عددی پیچیده با متغیرهای مرتبط چندگانه در مطالعات فراوانی دیده شده است. به طور خاص، محاسبات نرم و مدل های یادگیری ماشین در تجزیه و تحلیل چند متغیره در تحقیقات مهندسی آب و محیط زیست به خوبی عمل می کنند (Ebtehaj et al., 2015; Yaseen, 2021; Zayeri, 2023; Zhang et al., 2018). در مطالعه ای Yaseen (2023) به بررسی ضریب دبی عبوری سرریزهای قطاعی با استفاده از مدل های یادگیری ماشین پرداختند. نتایج آنان نشان میدهد با استفاده از مدل حافظه طولانی کوتاه مدت دو طرفه (Bi-LSTM) با حداقل مقدار میانگین درصد خطای مطلق ( $MAPE = 2.245$ ) و حداکثر شاخص ویلموت ( $WI = 0.861$ ) در مرحله اعتبار سنجی نسبت به سایر مدل ها ضریب دبی را به خوبی پیش بینی نموده است. (Salmasi et al. (2022) به تخمین میزان ضریب دبی در سرریزهای کشویی با استفاده از مدل های درخت تصمیم، ماشین بردار پشتیبان پرداختند. نتایج آنان نشان می دهد تحت شرایط جریان مستغرق مدل درخت تصمیم با شاخص های آماری ( $NSE = 0.954$ ,  $RMSE = 0.029$ ) کارآمد تر از مدل دیگر بوده است. Yan et al. (2023) در مطالعه ای به بررسی ضریب دبی عبوری

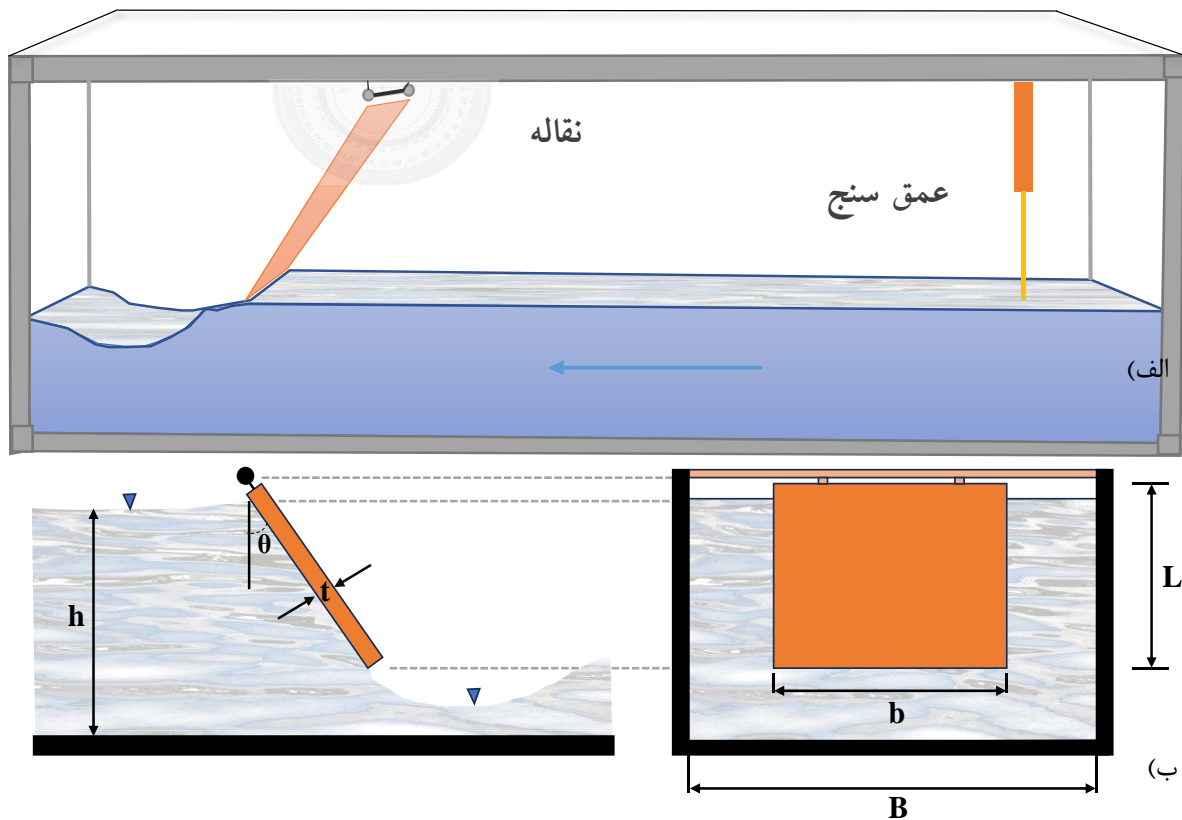
نظر گرفته شد که فرم بدون بعد آن در رابطه (۲) بیان شده است.

$$Q = \phi[h, B, b, S, t, L, g, M, \theta] \quad (۱)$$

$$\frac{Q}{\sqrt{gB^5}} = \phi\left(\frac{h}{B}, \frac{b}{B}, S, \frac{L}{B}, \frac{t}{B}, \frac{M}{\rho B^3}, \theta\right) \quad (۲)$$

که در آن  $Q$  دبی جریان (مترمکعب برثانیه)،  $B$  عرض فلوم (متر)،  $h$  عمق جریان بالادست (متر)،  $S$  شیب کف کانال،  $b$  عرض،  $L$  طول،  $t$  ضخامت صفحه آویخته،  $g$  شتاب ثقل و  $\theta$  زاویه بازشدگی صفحه آویخته می‌باشند که در شکل (۱) ب نشان داده شده است. مشخصات آماری داده‌های بدون بعد جمع‌آوری شده در جدول (۱) ارائه شده است.

(2022) Vatankhah که در آزمایشگاه هیدرولیک گروه مهندسی آبیاری دانشگاه تهران صورت پذیرفت، جهت مدل سازی استفاده گردید. نمایی از کانال اندازه‌گیری دبی جریان با دریچه آویخته در شکل ۱ الف نشان داده شده است. نتایج آزمایشگاهی مورد استفاده در این تحقیق در فلومی با عرض ثابت ۰/۲۵ متر ارتفاع ۰/۵ متر و طول ۱۲ متر برای دو شیب کف کانال  $S=0.005$  و افقی ( $S=0$ ) صورت پذیرفت. از آنجایی که دبی جریان عبوری از دریچه آویخته به مشخصات هندسی کانال و دریچه و مشخصات هیدرولیکی عمق بالادست جریان مرتبط است. از این رو در مطالعه حاضر برای توسعه مدل‌های SVM، GMDH و RF، هفت پارامتر ورودی بدون بعد در رابطه (۱) به دست آمده از آنالیز ابعادی مطابق با تحقیقات آزمایشگاهی در



شکل (۱): الف) شمای از کانال اندازه‌گیری دبی جریان با دریچه آویخته و ب) پارامترهای هندسی و هیدرولیکی موثر بر دبی

جدول (۱): آماره‌های توصیفی داده‌های تخصیص داده‌شده در مرحله آموزش و اعتبارسنجی مدل‌های محاسبات نرم

مرحله	شاخص	$\frac{Q}{\sqrt{gB^3}}$	$\frac{M}{\rho b^3}$	$\frac{L}{B}$	$\frac{t}{B}$	$\frac{b}{B}$	$\frac{h}{B}$	$\theta$	$S$
-------	------	-------------------------	----------------------	---------------	---------------	---------------	---------------	----------	-----

۰	۳/۵	۰/۰۶۵	۰/۲	۰/۰۲	۰/۸	۰/۰۲۹	۰/۰۱۱	کمینه	آموزش
۰/۰۰۵	۸۰	۰/۹۵۱	۰/۸	۰/۰۴	۱/۲	۱/۴۲۸	۰/۷۶	بیشینه	
۰/۰۰۲	۴۲/۱۱۱	۰/۴۰۸	۰/۴۹۹	۰/۰۳۰	۱/۰۰۹	۰/۳۲۹	۰/۲۳۶	میانگین	
۰/۰۰۲	۲۰/۰۹۳	۰/۲۲۱	۰/۲۲۱	۰/۰۱	۰/۲	۰/۴۰۰	۰/۱۹۳	انحراف معیار	
۰	۳/۵	۰/۰۷۲	۰/۸	۰/۰۲	۰/۸	۰/۰۲۹	۰/۰۱۱	کمینه	اعتبارسنجی
۰/۰۰۵	۸۱	۰/۸۸۷	۰/۲	۰/۰۴	۱/۲	۱/۴۲۸	۰/۷۳۶	بیشینه	
۰/۰۰۲	۴۲/۴۰۱	۰/۴۰۶	۰/۵۱۳	۰/۰۲۸	۱/۰۱۶	۰/۲۸۹	۰/۲۳۷	میانگین	
۰/۰۰۲	۲۰/۵۱۰	۰/۲۲۲	۰/۲۳۲	۰/۰۱	۰/۲۰۱	۰/۳۳۹	۰/۱۹۵	انحراف معیار	

خروجی است. این الگوریتم به لحاظ ساختاری شبیه شبکه‌های عصبی مصنوعی چندلایه و الگوریتم‌های داده‌کاوی است با این تفاوت که تعداد لایه‌ها و نرون‌ها توسط معیاری از پیش تعیین شده، مشخص می‌شود. الگوریتم GMDH در حل مسائل مختلف مهندسی هیدرولیک به طور گسترده‌ای استفاده شده است. از مهم‌ترین کاربردهای این روش می‌توان به برآورد آبشستگی اطراف پایه پل، پایین‌دست پرتاب‌کننده جامی شکل، ضریب دبی سازه‌های اندازه‌گیری جریان مانند انواع سرریزها اشاره نمود (Mahmoudi and Farhoudi, 2019; Parsaie et al., 2019). در واقع، الگوریتم GMDH برای ارزیابی رفتارهای محیطی سیستم‌های غیرخطی با پیچیدگی بالا به کار برده می‌شود. ایواخنکو با استفاده از چندجمله‌ای کولموگروف-گابور، تئوری GMDH را گسترش داد. ارتباط بین پارامترهای ورودی و خروجی هر سیستم را می‌توان با سری توابع ولترا که مشابه با چندجمله‌ای گسسته شده کولموگروف-گابور است، به صورت رابطه (۳) بیان کرد.

$$y = a_0 + \sum_{i=1}^m a_i x_i + \sum_{i=1}^m \sum_{j=1}^m a_{ijk} x_i x_j + \sum_{i=1}^m \sum_{j=1}^m \sum_{k=1}^m a_{ijk} x_i x_j x_k + \dots \quad (3)$$

که در آن  $X = (x_1, x_2, \dots, x_m)$  و  $A = (a_0, a_1, a_2, \dots, a_m)$  به ترتیب بردارهای پارامترهای ورودی و ضرایب وزنی هستند. او با استفاده از عملکرد شبکه‌های عصبی پرسپترون چندلایه، یک چندجمله‌ای درجه دو برای هر زوج پارامتر ورودی پیشنهاد داد.

در مطالعه آزمایشگاهی عمق جریان بالادست دریچه در خط مرکزی کانال با استفاده از یک نقطه سنج با دقت ۰/۱ میلی متر در فاصله ۰/۸ متری بالادست دریچه آویخته نصب شده بود اندازه گیری شد. صفحات از ورق پلکسی گلاس به صورت عمودی به گونه ای نصب می شدند که لبه های پایینی آنها در فاصله ۵ میلی متری از کف کانال قرار می گرفت. صفحات در فاصله ۸ متری از ورودی کانال نصب شدند. یک میله فولادی افقی به قطر ۵ میلی متر به عنوان محور لولا استفاده شد. زاویه انحراف،  $\theta$ ، بین محور عمودی و صفحات با استفاده از یک نقاله اندازه گیری می شد. در مجموع تعداد ۶۸۷ سری داده‌های آزمایشگاهی (۳۹۳ اجرا برای  $S=0$ ، و ۲۹۴ اجرا برای  $S=0/05$ ) مورد استفاده قرار گرفت. به منظور توسعه مدل‌های مبتنی بر محاسبات نرم لازم است که ابتدا داده‌های جمع‌آوری شده به دو دسته آموزش و اعتبارسنجی تقسیم‌بندی شوند. قابل ذکر است که در این تحقیق جهت صحت سنجی مدل های مورد استفاده پس از بررسی حالت های مختلف در نهایت ۸۰ درصد داده‌ها (۵۵۰ نمونه) به مرحله آموزش و ۲۰ درصد مابقی (۱۳۷ نمونه) به اعتبارسنجی مدل‌های توسعه یافته تخصیص داده شد. با توجه به این اینکه داده‌های جمع‌آوری شده ماهیت سری زمانی ندارند لذا اختصاص داده، به هریک از گروه‌های آموزش و اعتبارسنجی به صورت تصادفی انجام گرفت.

### مدل دسته‌بندی گروهی داده‌ها

الگوریتم دسته‌بندی گروهی داده‌ها (GMDH) اولین بار توسط ایواخنکو (Ivakhnenko, 1971) جهت تحلیل سیستم‌ها با درجات پیچیدگی بالا ارائه شد. مدل پیشنهادی ایواخنکو دارای لایه‌های ورودی، میانی و لایه

ساخت شبکه GMDH، ابتدا ترکیب‌های دوتایی از پارامترهای ورودی را باید در نظر گرفت. بعد از آن، در هر نرون با استفاده از روش حداقل مربعات خطا، ضرایب وزنی هر نرون و مقدار خطای مربوطه محاسبه می‌شوند. در هر لایه برای انتخاب نرون‌های برتر از نظر ویژگی طبقه‌بندی، از معیار زیر در رابطه (۵) استفاده می‌شود.

$$E = \frac{1}{n} \sum_{i=1}^{i=n} (y_{Observed(i)} - y_{Predicted})^2 \quad (5)$$

### مدل ماشین بردار پشتیبان

ماشین بردار پشتیبان (SVM) یک سیستم یادگیری کارآمد بر مبنای تئوری بهینه‌سازی مقید است که از اصل استقرای کمینه‌سازی خطای ساختاری استفاده کرده و منجر به یک جواب بهینه کلی می‌شود. در یک مدل رگرسیونی SVM لازم است وابستگی متغیر هدف (y) به مجموعه‌ای از متغیرهای مستقل (x) تخمین زده شود. فرض بر این است که مانند دیگر مدل‌های رگرسیونی، رابطه بین متغیرهای هدف و مستقل توسط یک تابع معین f به علاوه یک مقدار اضافی نویز<sup>۱</sup> به صورت رابطه (۶) مشخص می‌شود.

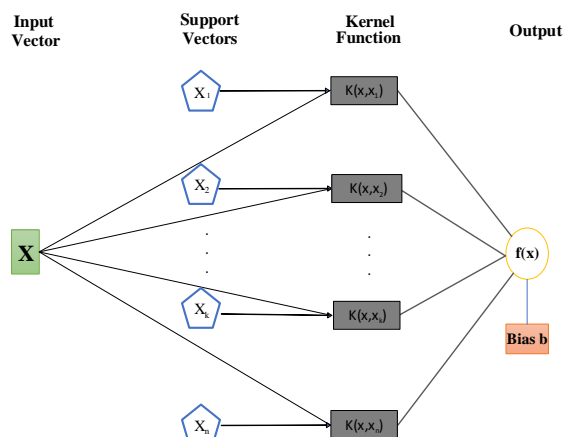
$$y = f(x) + noise \quad (6)$$

بنابراین موضوع اصلی پیدا کردن فرم تابع f است که بتواند به صورت صحیح، موارد جدیدی را که SVM تاکنون تجربه نکرده است پیش‌بینی کند. در روش SVM-ε از توابع هسته‌ای (کرنل) به منظور جداسازی صفحات و بهینه‌سازی فواصل موجود بین داده‌های مشاهداتی و تخمینی استفاده می‌شود، اما در مدل V-SVM از روش برنامه‌ریزی غیرخطی برای کاهش خطای پیش‌بینی استفاده می‌شود (Roushangar et al., 2017). در این مطالعه مدل SVM به دلیل کاربرد گسترده آن در مسائل رگرسیونی استفاده گردید. برای این مدل، تابع خطا به صورت زیر تعریف می‌شود:

همچنین او عنوان کرد که یک چندجمله‌ای درجه دوم می‌تواند در طی عملکرد شبکه‌های پرسپترون باعث تشکیل چندجمله‌ای کولموگروف-گابور شود. این روش مقایسه با ساختار نوع پرسپترون دارای دقت بیشتری است زیرا در الگوریتم GMDH محاسبات انجام‌شده در هر نرون به صورت داده‌های مفید و غیرمفید طبقه‌بندی می‌شوند. ساختار GMDH در قالب یک شبکه جلو رونده چندلایه از تعدادی نرون‌های پشتیبان ایجاد می‌شود. هر نرون حداقل دارای دو ورودی است. ارتباط بین متغیرهای ورودی و خروجی در هر نرون با استفاده از تابع محرک به صورت یک چندجمله‌ای خطی یا غیرخطی می‌تواند باشد، که به صورت رابطه (۴) بیان می‌شود.

$$y = O(x_1, x_2) = a_0 + a_1x_1 + a_2x_2 + a_3x_1^2 + a_4x_2^2 + a_5x_1x_2 \quad (4)$$

$a_0$  تا  $a_5$  ضرایب چندجمله‌ای هستند. در اغلب مطالعات مهندسی از رابطه (۴) جهت مدل‌سازی پدیده‌ها استفاده می‌شود. الگوریتم GMDH بر اساس اصل تکامل توسعه یافته است که در آن فرآیندهایی نظیر تولید نسل و انتخاب نرون‌های برتر در هر لایه دخالت دارند. در اولین لایه، نرون‌ها توسط ترکیب پارامترهای ورودی ایجاد می‌شوند و در لایه دوم آن‌ها به نرون‌های برتر قبلی متصل می‌شوند. انتخاب نرون‌های برتر که دارای بهترین ویژگی می‌باشند بر اساس معیارهای خارجی که می‌توانند خطای موجود در هر نرون را ارزیابی کنند، صورت می‌گیرد. در مدل GMDH برای ایجاد چندجمله‌ای‌های درجه دو در نرون‌های لایه اول، ترکیب‌های دوتایی پارامترهای ورودی در نظر گرفته می‌شوند. اگر تعداد شبکه GMDH دارای P پارامتر ورودی باشد، تعداد نرون‌های لایه اول برابر با  $L = \binom{P}{2}$  است. همچنین برای تولید نرون‌های لایه‌های میانی، ترکیب‌های دوتایی نرون‌های برتر در لایه قبلی در نظر گرفته می‌شوند، به عبارت دیگر مدل GMDH نوعی از شبکه‌های چندلایه جلو رونده است که به صورت ترکیبی از دسته معادلات چندجمله‌ای درجه دو بیان می‌شوند. برای



شکل (۲): ساختار مدل ماشین بردار پشتیبان

پارامتر  $K(x, x_i)$  تابع کرنلی است که برای ایجاد ماشین‌هایی با انواع مختلف از سطوح غیرخطی در فضای داده‌ها، ضرب‌های داخلی تولید می‌کند. برای مدل رگرسیون ماشین بردار پشتیبان از کرنل‌های مختلفی استفاده می‌شود که عبارت‌اند از خطی، درجه دوم، گوسی و چندجمله‌ای. معمولاً تابع کرنل گوسی شعاعی (RBF) و یا خطی برای پیش‌بینی عملکرد بهتری دارند (Zayeri, 2023). رابطه تابع کرنل RBF به صورت رابطه (۱۳) بیان می‌شود.

$$K(x, y) = \exp\left(-\frac{\|x_i - y_j\|}{2\sigma^2}\right) \quad (13)$$

### مدل جنگل تصادفی

مدل جنگل تصادفی (RF) طبقه‌بندی را برای بسیاری از مجموعه داده‌ها، با سرعت بالایی انجام می‌دهد و بر خلاف مدل‌های نظیر رگرسیون چندمتغیره با استفاده از صدها و هزاران درخت از اطلاعات بیشتر موجود در داده‌ها استفاده می‌کند تا بتواند بهترین استنباط از متغیرهای ورودی داشته باشد. این روش یک روش متداول از مدل‌های نلپارامتری بوده که از تکنیک بگینگ<sup>۲</sup> استفاده می‌نماید که ترکیبی از چندین درخت تصمیم است که در ساخت آن چندین نمونه بوت استرپ از داده‌ها شرکت دارند و در ساخت هر درخت به طور تصادفی تعدادی از متغیرهای ورودی شرکت می‌کنند از این رو برخی از داده‌ها ممکن است بیش از یکبار در شاخه‌ای آموزشی استفاده

$$\frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^n (\xi_i^- + \xi_i^+) \quad (7)$$

تابع خطای مذکور لازم است که با توجه به محدودیت‌های زیر حداقل گردد:

$$y_i - (\langle \omega, x_i \rangle + b) \leq \varepsilon + \xi_i \quad (8)$$

$$(\langle \omega, x_i \rangle + b) - b \leq \varepsilon + \xi_i^* \quad (9)$$

$$\xi_i, \xi_i^* \geq 0 \quad (10)$$

که در این روابط C یک مقدار از پیش تعیین شده می‌باشد،  $\omega$  بردار ضرایب،  $\xi_i^-$  و  $\xi_i^+$  متغیرهایی هستند که قیود بالا و پایین خروجی سیستم را تعیین می‌کنند. اگر داده‌ها به صورت خطی مجزا از هم باشند،  $\xi_i^-$  و  $\xi_i^+$  یک سطح بهینه که داده‌ها را بدون خطا و با حداکثر فاصله میان صفحه و نزدیک‌ترین نقاط آموزشی (بردارهای پشتیبان) تفکیک می‌نماید، آموزش می‌دهد. مطابق شکل (۲) اگر نقاط آموزشی را به صورت  $[x_i, y_i]$  و بردار ورودی  $x_i \in R^n$  تعریف شود و در حالتی که داده‌ها به صورت خطی قابل تفکیک باشند، رابطه (۱۱) را به صورت زیر می‌توان بیان نمود.

$$y = f(x) = \text{sign}\left[\sum_{i=1}^N y_i a_i \langle x_i, x \rangle + b\right] \quad (11)$$

که در آن  $y$  خروجی معادله و  $y_i$  ارزش طبقه نمونه آزمایشی  $x_i$  است. بردار  $x = (x_1, x_2, \dots, x_n)$  نشان دهنده یک داده ورودی و بردارهای  $x_i, i=1, 2, \dots, N$  بردارهای پشتیبان هستند. اگر داده‌ها به صورت خطی قابل تفکیک نباشند، می‌توان با اعمال پیش پردازش، نمونه‌ها را به فضایی بالاتر برد. در این حالت رابطه (۱۱) به (۱۲) تغییر می‌یابد.

$$y = f(x) = \text{sign}\left[\sum_{i=1}^N y_i a_i K(x, x_i) + b\right] \quad (12)$$

$$R^2 = \frac{\sum_{i=1}^N (O_i - \bar{O}_i)(P_i - \bar{P}_i)}{\sum_{i=1}^N (O_i - \bar{O}_i)^2 + \sum_{i=1}^N (P_i - \bar{P}_i)^2} \quad (14)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (O_i - P_i)^2}{N}} \quad (15)$$

در رابطه های (۱۴) و (۱۵)،  $O_i$  مقادیر دبی مشاهداتی (واقعی)،  $P_i$  مقادیر دبی پیش‌بینی شده،  $\bar{O}_i$  میانگین مقادیر دبی مشاهداتی،  $\bar{P}_i$  میانگین مقادیر دبی پیش‌بینی شده و  $N$  تعداد مشاهدات هستند. لازم به ذکر است مقادیر  $R^2$  نزدیک یک و مقادیر  $RMSE$  نزدیک به صفر، نشان‌دهنده دقت بیشتر و عملکرد بهتر مدل هستند. در این مطالعه، علاوه بر روش‌های ارزیابی اشاره شده از نمودار تیلور (Taylor, 2001) به منظور ارزیابی عملکرد مدل‌ها نیز استفاده شده است. این نمودار بر مبنای ارتباط هندسی بین ضریب همبستگی، انحراف معیار و انحراف جذر میانگین مربعات  $(RMSD)$  استوار است. در نمودار تیلور که به دو صورت نیم دایره (نمایش همبستگی منفی و مثبت) و ربع دایره (فقط نمایش همبستگی مثبت) ارائه می‌شود، مقادیر ضریب همبستگی به صورت شعاع دایره روی قوس آن و مقادیر انحراف معیار به صورت دایره های متحدالمرکز نسبت به مرکز دایره و مقادیر  $RMSD$  به صورت دایره های متحدالمرکز نسبت به نقطه مرجع ترسیم میشوند. نقطه مرجع موقعیت داده دبی مشاهده ای را بر اساس انحراف معیار آن نشان می‌دهد و موقعیت آن روی محور افقی بر اساس مقدار انحراف معیار تعیین خواهد شد زیرا مقدار  $RMSD$  و ضریب همبستگی دبی مشاهده ای در مقایسه با خودش به ترتیب صفر و یک خواهد بود. بنابراین، روش ارزیابی در این نمودار به این صورت است که موقعیت داده‌های مورد بررسی بر اساس مقدار  $RMSD$  و ضریب همبستگی آن با داده‌های مشاهده ای و انحراف معیار آن روی نمودار ترسیم می‌شود و هر داده ای که موقعیت مکانی آن روی نمودار به نقطه مرجع

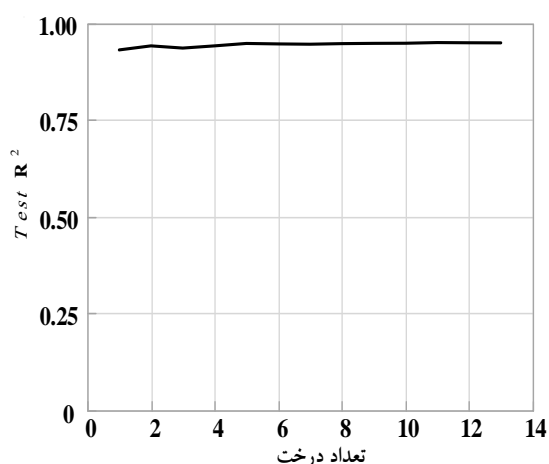
شوند در حالی که، برخی از داده‌های دیگر که در مدل‌سازی مؤثر نیستند هرگز استفاده نمی‌شوند. بنابراین مدل ثبات بیشتری به دست می‌آورد و در برابر تغییرات جزئی در داده‌های ورودی قابل اعتماد تر می‌کند و دقت پیش‌بینی آن را افزایش می‌دهد. از سوی دیگر، هنگامی که در مدل RF یک درخت ایجاد می‌شود از بهترین متغیرها یا نقاط تقسیم در داخل زیرمجموعه‌های متغیرها استفاده می‌کند که به صورت تصادفی از مجموعه‌های کلی متغیرهای ورودی انتخاب می‌شوند. لذا در مدل RF با استفاده از روش بوت استرپ به تعداد زیاد (به عنوان مثال ۱۰۰۰ مرتبه) نمونه‌های  $n$  تایی از مجموعه داده‌های اولیه ورودی، نمونه برداری همراه با جای‌گذاری انجام می‌شود. در طی فرآیند نمونه‌گیری حدود یک سوم از داده‌ها به عنوان نمونه خارج از نمونه‌گیری (از این داده‌ها برای تعیین متغیرهای مهم و همچنین برآورد نا اریب خطا استفاده می‌شود) در نظر گرفته می‌شوند و نمونه‌گیری از آنها صورت نمی‌پذیرد. سپس بر روی هر نمونه بوت استرپ یک درخت تصمیم گسترش داده می‌شود. در طی فرآیند ساخت درخت در هر شاخه، از بین تمام  $M$  متغیر مستقل به صورت تصادفی  $m$  متغیر برای تقسیم شدن انتخاب می‌شود. برای حالت رگرسیونی نسبت  $Mm$  برابر با یک سوم است و برای کلاسه بندی برابر با  $m = \sqrt{M}$  پیشنهاد شده است پس از ساخت تمام درخت داده‌های آزمایش به درخت معرفی شده و به تعداد درخت‌ها برای بردار ورودی یک خروجی به دست می‌آید. با میانگین‌گیری این خروجی‌ها، خروجی نهایی مدل و با در نظر گرفتن توزیع تجربی خروجی‌ها مقادیر صدک‌ها و دامنه عدم قطعیت محاسبه می‌شود. روش درخت رگرسیون جنگل تصادفی به ویژه هنگامی که تعداد مشاهدات در مقایسه با تعداد پیش‌بینی‌کننده‌ها نسبتاً کم باشد یک روش پیش‌بینی کارآمد است (Siasar and Honar, 2019).

### معیارهای ارزیابی

به منظور ارزیابی مدل‌های مورد استفاده در این تحقیق از شاخص‌های آماری ضریب تعیین، ریشه میانگین مربعات خطا به شرح زیر استفاده گردید:

$$0.176 \frac{Q}{\sqrt{gB^5}} = 0.979 \frac{h}{B} - 0.187 \quad 60 < \theta \leq 80$$

در ابتدا نتایج حاصل از مدل جنگل تصادفی ارائه می‌شود. دلیل اولویت ارائه نتایج مدل جنگل تصادفی نسبت به سایر مدل‌های بکار رفته در این تحقیق مشخص شدن مهم‌ترین پارامترهای مؤثر در فرایند توسعه مدل جنگل تصادفی در مدل‌سازی و برآورد دبی جریان است. در مدل GMDH نیز همین ویژگی دیده می‌شود. در شکل (۴) و (۵) نتایج مدل جنگل تصادفی در مراحل آموزش و اعتبارسنجی در مقابل مقادیر مشاهداتی ترسیم شده است. شاخص‌های آماری مدل جنگل تصادفی توسعه داده شده در مرحله آزمایش  $R^2 = 0.967$ ،  $RMSE = 0.035$  و در مرحله اعتبارسنجی عبارت از  $R^2 = 0.955$ ،  $RMSE = 0.041$  است. مطابق شکل (۶) تعداد ۱۳ درخت تصمیم استفاده شده در برآورد مقدار دبی عبوری از دریاچه آویخته استفاده گردید با این حال از درخت سوم مقدار ضریب تعیین  $R^2$  تغییر محسوسی نداشته است لذا مدل با میانگین سه درخت تصمیم توسعه یافته اول می‌تواند پیش بینی مناسبی ارائه دهد. برای نمونه نتایج بهترین مدل درخت تصمیم در مرحله اعتبارسنجی مدل RF در شکل (۷) نشان داده شده است.

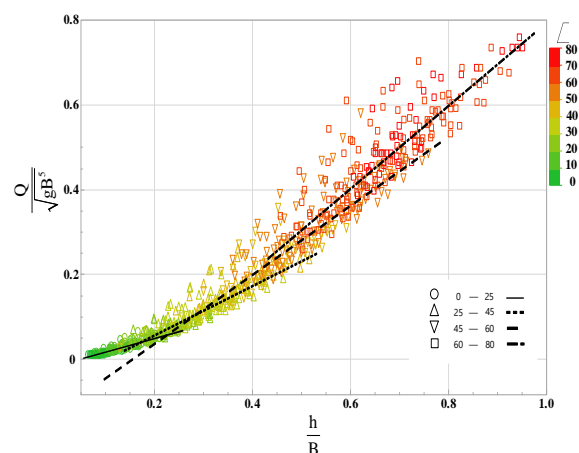


شکل (۶): تأثیر افزایش مدل درخت تصمیم در ساختار RF بر میزان برآورد دبی در مرحله اعتبارسنجی

نزدیکتر باشد، بادقت بیشتری مقادیر دبی را شبیه‌سازی می‌نماید.

## نتایج و بحث

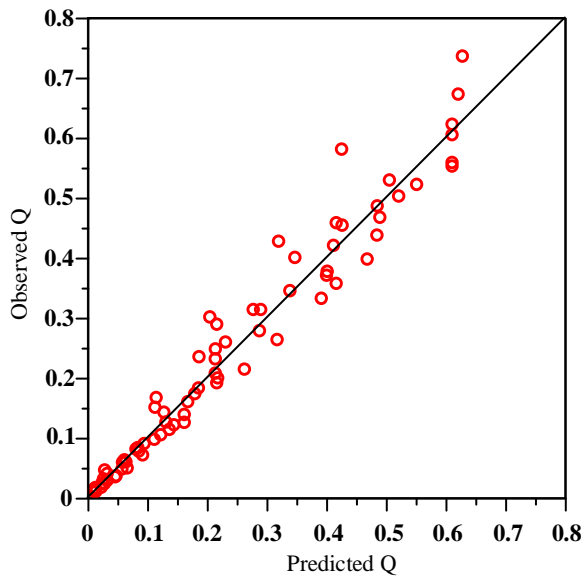
به منظور ارزیابی نتایج آزمایشگاهی، پارامتر بدون بعد دبی نسبت به عمق نسبی به ازای زوایای مختلف بازشدگی دریاچه آویخته در شکل (۳) نشان داده شده است. همانگونه که مشاهده می‌شود با افزایش عمق نسبی، میزان دبی و زاویه بازشدگی دریاچه افزایش یافته است. این روند در چهار محدوده زاویه بازشدگی صفر تا ۸۰ درجه به همراه روابط رگرسیون خطی در جدول (۲) نشان داده شده است. مطابق با جدول فوق با افزایش ۵ برابری زاویه بازشدگی دریاچه آویخته میزان دبی نسبی عبوری تا سه برابر افزایش داشته است.



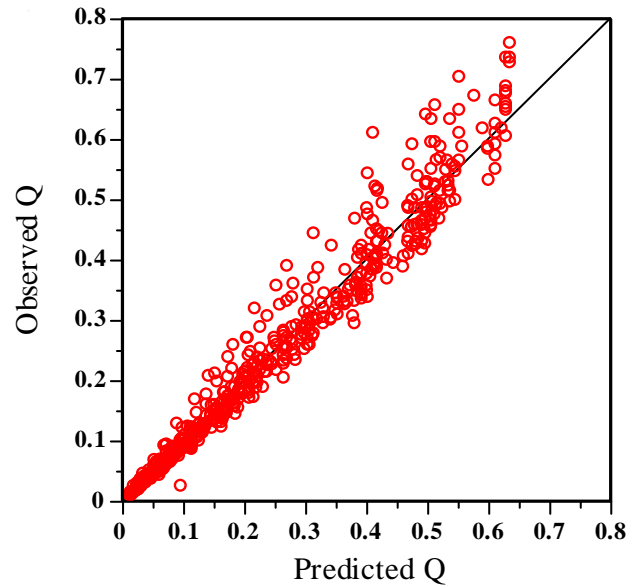
شکل (۳): تغییرات دبی عبوری از دریاچه آویخته با عمق جریان نسبی و زوایای بازشدگی مختلف

جدول (۲): روابط رگرسیون خطی دبی به ازای زوایای بازشدگی دریاچه آویخته

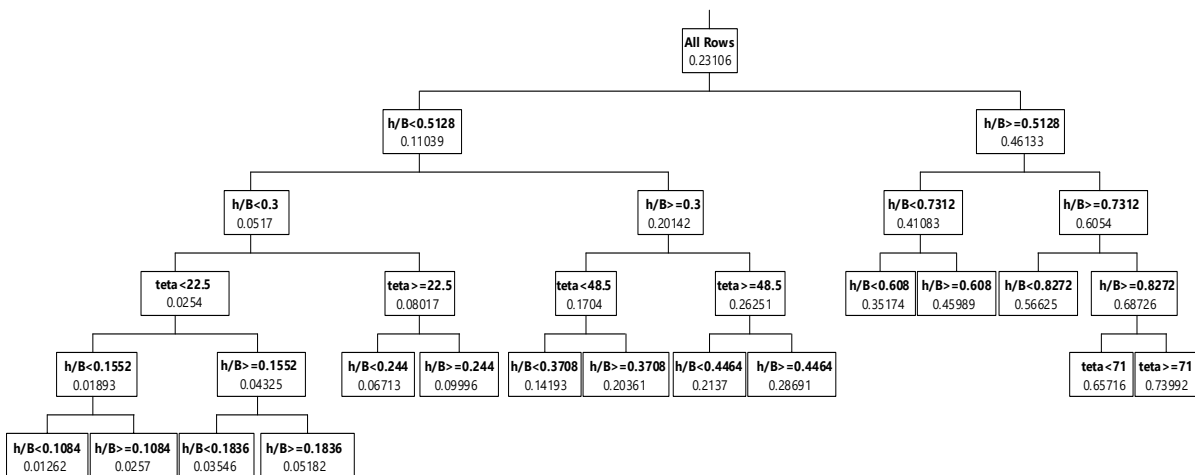
$R^2$	رگرسیون خطی	محدوده بازشدگی دریاچه
۰.۱۸۵	$\frac{Q}{\sqrt{gB^5}} = 0.322 \frac{h}{B}$	$0 < \theta \leq 25$
۰.۱۷۷	$\frac{Q}{\sqrt{gB^5}} = 0.583 \frac{h}{B} - 0.062$	$25 < \theta \leq 45$
۰.۱۷۴	$\frac{Q}{\sqrt{gB^5}} = 0.815 \frac{h}{B} - 0.128$	$45 < \theta \leq 60$



شکل (۵): نتایج مدل RF توسعه داده شده در مرحله اعتبارسنجی



شکل (۴): نتایج مدل RF توسعه داده شده در مرحله آموزش



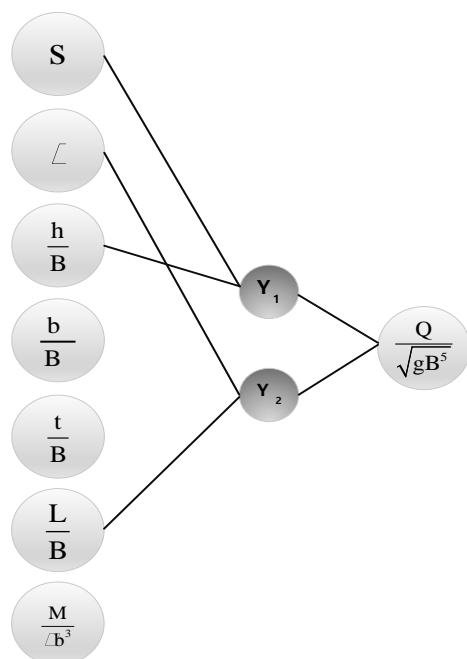
شکل (۷): ساختار بهترین درخت تصمیم در مدل RF جهت برآورد دبی جریان

$$R_{ij} = \frac{\sum_{k=1}^N I_{ik} O_{jk}}{\sqrt{\sum_{k=1}^N I_{ik}^2 \cdot \sum_{j=1}^M O_{jk}^2}} \quad (16)$$

که در آن  $I_i$  و  $O_j$  به ترتیب متغیرهای ورودی و خروجی هدف، در رابطه (۲) می توان با جابجایی هر متغیر ورودی با نرخ ثابت و ثابت نگه داشتن متغیرهای ورودی دیگر، مطابق با رابطه (۱۶)، از روش دامنه کسینوسی بهره برد. (Momeni et al., 2014)

### تحلیل حساسیت داده ها

برای نشان دادن تأثیر هر متغیر ورودی بر متغیر هدف، در رابطه (۲) می توان با جابجایی هر متغیر ورودی با نرخ ثابت و ثابت نگه داشتن متغیرهای ورودی دیگر، مطابق با رابطه (۱۶)، از روش دامنه کسینوسی بهره برد. (Momeni et al., 2014)



شکل (۹): ساختار نهایی مدل GMDH در پیش‌بینی میزان دبی نسبی

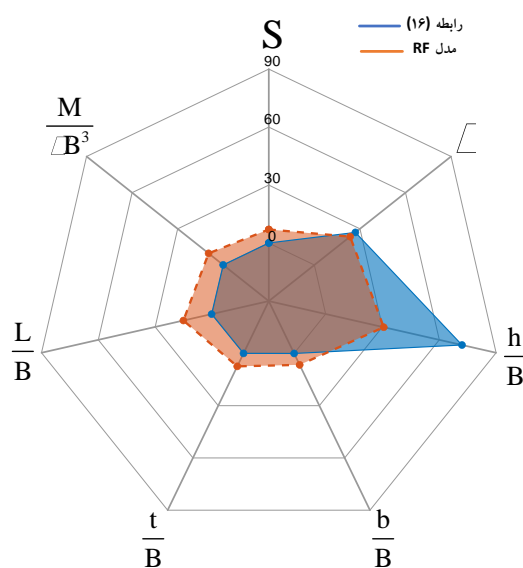
شکل (۱۰) و (۱۱) نتایج مدل GMDH در مراحل آموزش و اعتبارسنجی در مقابل مقادیر مشاهداتی ترسیم شده است. شاخص‌های آماری مدل GMDH توسعه داده شده در مرحله آزمایش عبارت از  $RMSE = 0.023$  و در مرحله اعتبارسنجی عبارت از  $R^2 = 0.984$  و  $RMSE = 0.0981$  است.

در روابط (۱۷) تا (۱۹) معادلات به دست آمده از نروون‌های لایه پنهان در برآورد میزان دبی خروجی در مدل GMDH نشان داده شده است. در شکل (۱۲) هیستوگرام میزان خطای نسبی رابطه (۱۷) در تخمین میزان دبی عبوری نشان داده شده است. مطابق با این نمودار، رابطه پیشنهادی مدل GMDH با  $\pm 5\%$  درصد خطا در محدوده سه برابر انحراف معیار از میانگین صفر قرار گرفته که در شکل (۱۳) نتایج حاصل از کل مشاهدات در مقابل مدل GMDH حاصل از رابطه (۱۷) نشان داده شده است.

زاویه بازشدگی دریاچه بیشتر تاثیر و پارامتر  $\frac{M}{\rho b^3}$  و  $\frac{t}{B}$

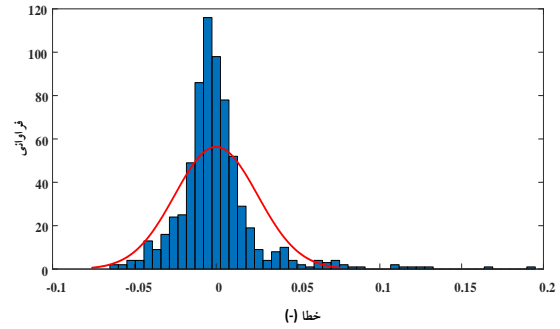
کمترین تاثیر را در برآورد مقدار  $\frac{Q}{\sqrt{gB^5}}$  داشته است.

در ادامه، عملکرد مدل GMDH در برآورد میزان دبی جریان مورد بررسی قرار گرفت. ساختار نهایی مدل توسعه داده شده در شکل (۹) نشان داده شده است.

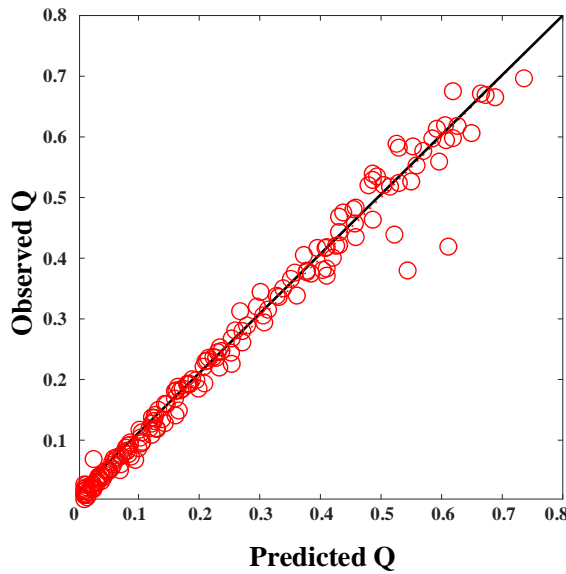


شکل (۸): میزان تأثیر مؤلفه‌های مختلف ورودی به مدل جنگل تصادفی در پیش‌بینی دبی عبوری از دریاچه آویخته

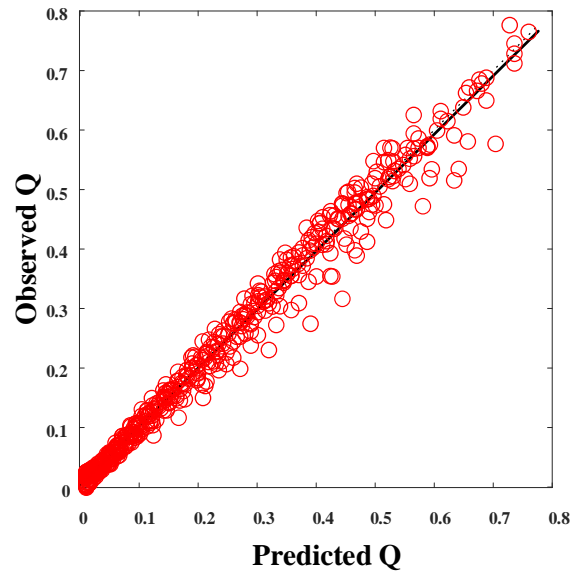
همان‌طور که مشاهده می‌شود، ساختار مدل GMDH دارای یک لایه پنهان است که دارای دو نرون و در لایه خروجی نیز یک نرون که متوسط جواب‌های دو نرون لایه قبل است وجود دارد. همان‌گونه که مشاهده می‌شود در ساختار نهایی به دست آمده از مدل GMDH، لایه خروجی  $\frac{Q}{\sqrt{gB^5}}$  فقط به دو نیال  $Y_1$  و  $Y_2$  در لایه پنهان متصل بوده که این دو مؤلفه نیز به متغیرهای  $S$ ،  $\theta$ ،  $h/B$  و  $L/B$  در لایه ورودی متصل بوده‌اند لذا مابقی متغیرهای ورودی در پیش‌بینی مقدار دبی نسبی تأثیرگذار نیستند. شایان ذکر است مقدار ریشه میانگین مربعات خطا در لایه اول برابر  $0.031$  و در لایه دوم  $0.026$  بوده است.



شکل (۱۲): هیستوگرام خطای نسبی مدل GMDH در تخمین میزان دبی عبوری از دریچه آویخته



شکل (۱۱): نتایج مدل GMDH توسعه داده شده در مرحله اعتبارسنجی

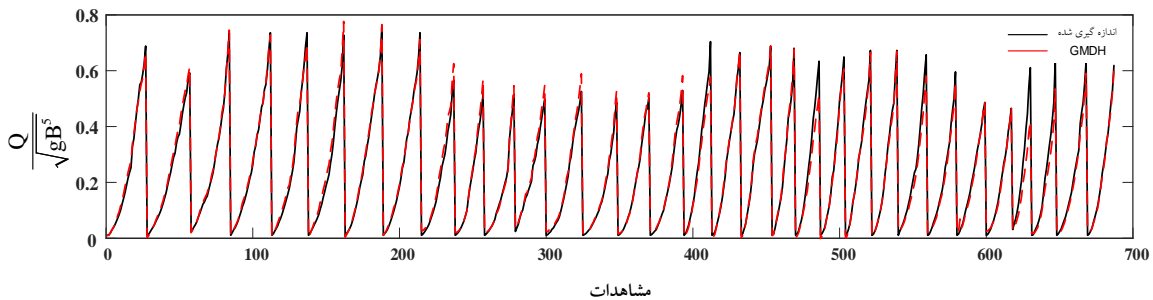


شکل (۱۰): نتایج مدل GMDH توسعه داده شده در مرحله آموزش

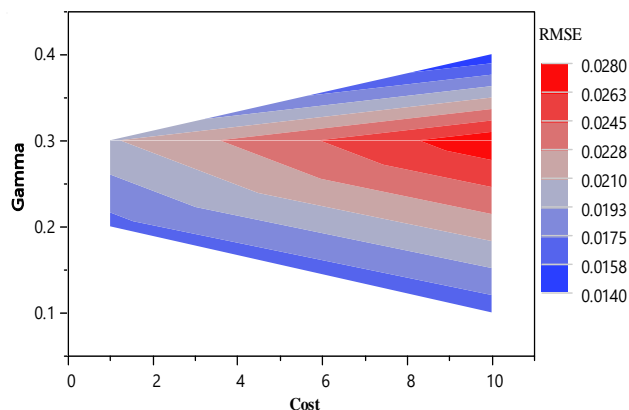
$$\frac{Q}{\sqrt{gB^5}} = -0.0057 + 0.446 \times Y_1 + 0.4803 \times Y_2 - 3.3644 \times Y_1^2 - 3.0463 \times Y_2^2 + 6.5861 \times Y_1 \times Y_2 \quad (17)$$

$$Y_1 = -0.0455 + 3.2355' S + 0.3813' \frac{h}{B} + 0.0162' S^2 + 0.5043' \left(\frac{h}{B}\right)^2 + 11.8829' S' \frac{h}{B} \quad (18)$$

$$Y_2 = 0.1137 - 0.0120 \times q + 0.0086 \times \frac{L}{B} + 1.3645 \times 10^{-4} \times (q)^2 - 0.0919 \times \left(\frac{L}{B}\right)^2 + 0.01 \times (q) \times \left(\frac{L}{B}\right) \quad (19)$$



شکل (۱۳): نتایج مدل GMDH در پیش‌بینی میزان دبی عبوری از دریچه آویخته در مقابل نتایج اندازه‌گیری شده

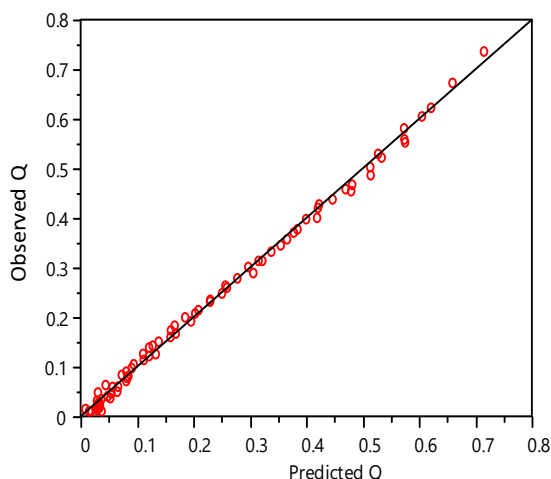


شکل (۱۴): تغییرات پارامترهای Cost و gamma مدل SVM در دقت برآورد میزان دبی جریان

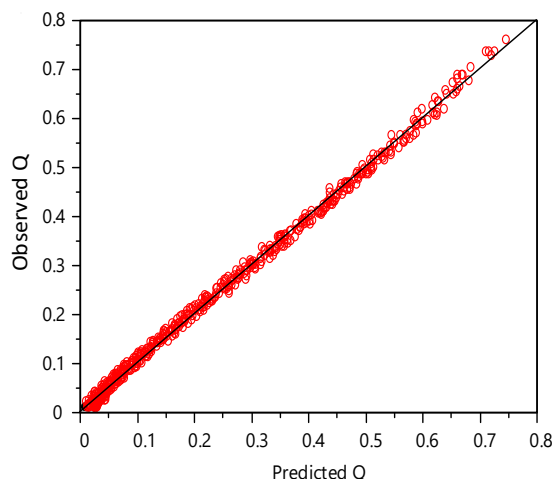
همچنین عملکرد مدل SVM جهت برآورد دبی جریان مورد بررسی قرار گرفت. به منظور توسعه مدل SVM توابع کرنل مختلفی که در قسمت مواد و روش‌های معرفی شده اند بررسی شدند و نتایج بیان شده داده در جدول (۳) نشان می‌دهد که تابع گوسی شعاعی با پارامترهای ضریب پنالتی (Gamma) برابر با ۰/۱ و پارامتر خطا (Cost) برابر با ۱۰ که در شکل (۱۴) آورده شده است به عنوان مدل چهارم دقت بهتری نسبت به سایر مدل‌های دیگر دارد. در شکل (۱۵) و (۱۶) به ترتیب به مقایسه نتایج به دست آمده از مدل SVM در مقابل نتایج دبی جریان اندازه‌گیری شده در مرحله آموزش و اعتبارسنجی پرداخته شده است. شاخص‌های آماری مدل SVM توسعه داده شده در مرحله آموزش عبارت از  $R^2=0.996$ ,  $RMSE=0.10$  و در مرحله اعتبارسنجی عبارت از  $R^2=0.996$  و  $RMSE=0.11$  است. مقایسه عملکرد مدل‌های یادگیری ماشین در پیش‌بینی میزان دبی نشان می‌دهد عملکرد مدل SVM از GMDH و مدل RF با ضریب تعیین ( $R^2$ ) بالاتر، بیشتر بوده است.

جدول (۲): بررسی توابع مختلف به عنوان کرنل در مدل SVM

مدل	تابع کرنل	Cost	Gamma	Training RMSE	Test	
					RMSE	$R^2$
Model 1	گوسی شعاعی	۱	۰/۲	۰/۰۱۳	۰/۰۱۶۸	۰/۹۹۲
Model 2	گوسی شعاعی	۱	۰/۳	۰/۰۱۴	۰/۰۲۰۸	۰/۹۸۸
Model 3	گوسی شعاعی	۱۰	۰/۴	۰/۰۱۳	۰/۰۱۶۳	۰/۹۹۲
<b>Model 4</b>	<b>گوسی شعاعی</b>	<b>۱۰</b>	<b>۰/۱</b>	<b>۰/۰۱۰</b>	<b>۰/۰۱۱</b>	<b>۰/۹۹۶</b>
Model 5	خطی	۱۰	-	۰/۰۳۲	۰/۰۲۷	۰/۹۷۹
Model 6	خطی	۲	-	۰/۰۳۲	۰/۰۲۷	۰/۹۷۹
Model 7	خطی	۱	-	۰/۰۳۲	۰/۰۲۷	۰/۹۷۹



شکل (۱۶): نتایج مدل SVM توسعه داده شده در مرحله اعتبارسنجی



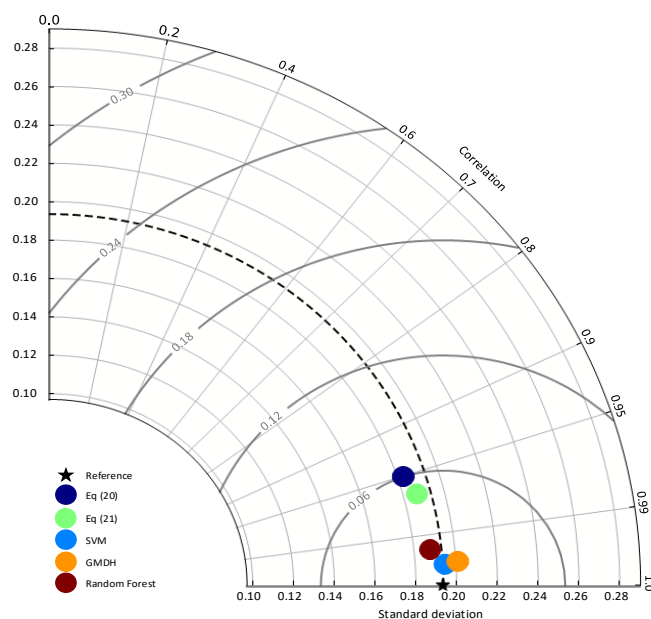
شکل (۱۵): نتایج مدل SVM توسعه داده شده در مرحله آموزش

ترسیم گردید. همان‌طور که مشاهده می‌شود عملکرد مدل SVM نسبت به هر دو مدل GMDH و RF و روابط (۲۰) و (۲۱) با RMSD کمتر و نزدیک تر بودن به نقطه مرجع مشاهداتی بسیار دقیق‌تر در پیش‌بینی میزان دبی خروجی از دریچه آویخته بوده است.

جهت بررسی بهتر عملکرد مدل‌های توسعه داده شده در این تحقیق، به مقایسه دو معادله رگرسیون غیرخطی ارائه شده در روابط (۲۰) و (۲۱) ارائه شده توسط گرگین و وطن‌خواه (Gorgin and Vatankhah, 2022) در پیش‌بینی میزان دبی عبوری از این نوع از دریچه‌ها صورت پذیرفت. دیاگرام تیلور برای نتایج آن‌ها در شکل (۱۷)

$$\frac{Q}{\sqrt{gB^5}} = (0.547 - 36S) \left(\frac{h}{B}\right)^{1.82-8S} \left(\frac{b}{B}\right)^{-0.13-40S} \left(\frac{t}{B}\right)^{-0.12-23.3S} \left(\frac{L}{B}\right)^{-0.1} \quad (20)$$

$$\frac{Q}{\sqrt{gB^5}} = (0.3 - 30S) \left(\frac{h}{B}\right)^{1.67-23S} \left(\frac{b}{B}\right)^{-0.13-34S} \left(\frac{t}{B}\right)^{-0.09-22.2S} \theta^{0.152+14S} \quad (21)$$



شکل (۱۷): دیاگرام تیلور عملکرد مدل‌های توسعه یافته

های  $R^2 = ۰/۹۹۶$ ,  $RMSE = ۰/۰۱۰$  در مرحله اعتبارسنجی تخمین دقیق تری در برآورد میزان دبی عبوری از دريچه آویخته داشته است. همچنین نتایج رگرسیون خطی نشان میدهد با افزایش ۵ برابری زاویه بازشدگی دريچه آویخته میزان دبی نسبی عبوری تا ۳ برابر افزایش می‌یابد. امید است که مطالعات تحقیقاتی آینده نیز در این راستا با بهبود دقت پیش‌بینی، به ویژه برای مقادیر محدوده بالاتر میزان دبی، با ترکیب یا بهبود در اندازه‌گیری بیشتر پارامترهای مدل متمرکز شوند.

## نتیجه‌گیری

در این تحقیق، تخمین دبی جریان در کانال‌های اندازه‌گیری دبی جریان با دريچه آویخته با استفاده از مدل‌های محاسبات نرم شامل مدل‌های SVM، GMDH، RF و مدل‌سازی گردید. نتایج این تحقیق نشان داد که در مقایسه با مدل رگرسیون غیر خطی ارائه شده توسط نتایج Gorgin and Vatankhah (2022) تمام مدل‌های محاسبات نرم توسعه یافته در مطالعه حاضر، عملکرد دقیق‌تر در مدل‌سازی برآورد میزان دبی در کانال با دريچه آویخته را نشان می‌دهد. بررسی ساختار مدل جنگل تصادفی در تحقیق حاضر نشان داد که مهمترین پارامترهای دخیل در برآورد دبی عبارت از نسبت عمق آب به عرض کانال ( $\frac{h}{B}$ ) به میزان ۷۲ درصد و کمترین مؤلفه تاثیرگذار بر میزان پیش‌بینی دبی عبوری از دريچه آویخته نسبت ( $\frac{t}{B}$ ) با مقدار ۰/۰۱ درصد بوده است. از میان مدل‌های یادگیری ماشین نامبرده، مدل ماشین بردار پشتیبان با شاخص



## سیاس‌گذاری

بدین وسیله از حمایت مالی معاونت پژوهش و فناوری دانشگاه شهید چمران اهواز در قالب پژوهانه (SCU.WH1403.43525) در انجام این تحقیق قدردانی می‌شود.

## منابع

- Burt, C.M., Angold, R., Lehmkuhl, M., Styles, S., 2001. Flap Gate Design for Automatic Upstream Canal Water Level Control. *J. Irrig. Drain. Eng.* 127, 84–91. [https://doi.org/10.1061/\(ASCE\)0733-9437\(2001\)127:2\(84\)](https://doi.org/10.1061/(ASCE)0733-9437(2001)127:2(84))
- Ebtehaj, I., Bonakdari, H., Zaji, A.H., Azimi, H., Sharifi, A., 2015. Gene expression programming to predict the discharge coefficient in rectangular side weirs. *Appl. Soft Comput.* 35, 618–628. <https://doi.org/10.1016/j.asoc.2015.07.003>
- Ghaderinia, H., Seyedzadeh, A., Vatankhah, A., 2019. Discharge Estimation in Semicircular Canals using Flap Plate. *Iran. J. Soil Water Res.* 50, 1183–1191. <https://doi.org/10.22059/ijswr.2018.261828.667965>
- Gorgin, F., Vatankhah, A.R., 2022. Rectangular top-hinged plate as portable flow measuring device. *Water Supply* 22, 8637–8658.
- Ivakhnenko, A.G., 1971. Polynomial Theory of Complex Systems. *IEEE Trans. Syst. Man Cybern.* SMC-1, 364–378. <https://doi.org/10.1109/TSMC.1971.4308320>
- J. A. Replogle, B. T. Wahlin, 2003. HEAD LOSS CHARACTERISTICS OF FLAP GATES AT THE ENDS OF DRAIN PIPES. *Trans. ASAE* 46. <https://doi.org/10.13031/2013.13970>
- Kartal, V., Emin Emiroglu, M., Katipoglu, O.M., 2024. Modeling of discharge capacity of H-weir using experiments, bio-inspired optimization and data preprocess based on SVM. *Int. J. Environ. Sci. Technol.* <https://doi.org/10.1007/s13762-024-05494-y>
- Litrico, X., Belaud, G., Baume, J.P., Ribot-Bruno, J., 2005. Hydraulic modelling of an automatic upstream water level control gate. *J. Irrig. Drain. Eng.* 131, 176–189. [https://doi.org/10.1016/\(ASE\)0733-9437\(2005\)131:2\(176\)](https://doi.org/10.1016/(ASE)0733-9437(2005)131:2(176))
- Mahmoudi, B., Farhoudi, J., 2019. Discharge estimation of flap gate installed in a circular channel under free flow condition. *Iran. Water Res. J.* 13, 127–136.
- Moghtaderi, A., Valizadegan, E., 2021. Investigation of the Hydraulic Characteristics of Flap Gates at the End of Trapezoidal Channels. *JWSS - Isfahan Univ. Technol.* 25, 107–118. <https://doi.org/10.47176/jwss.25.2.42921>
- Momeni, E., Nazir, R., Jahed Armaghani, D., Maizir, H., 2014. Prediction of pile bearing capacity using a hybrid genetic algorithm-based ANN. *Measurement* 57, 122–131. <https://doi.org/10.1016/j.measurement.2014.08.007>
- Parsaie, A., Haghiaibi, A.H., Emamgholizadeh, S., Azamathulla, H.M., 2019. Prediction of discharge coefficient of combined weir-gate using ANN, ANFIS and SVM. *Int. J. Hydrol. Sci. Technol.* 9, 412. <https://doi.org/10.1504/IJHST.2019.102422>
- Rabiee Moghadam, A., Khodashenas, S., Ziaie, A.N., 2018. Hydraulic study and Design of modified flap gate to automate the water level control in canals. *Iran. J. Irrig. Drain.* 12, 335–343.
- Roushangar, K., Saghebian, S.M., Mouaze, D., 2017. Predicting characteristics of dune bedforms using PSO-LSSVM. *Int. J. Sediment Res.* 32, 515–526. <https://doi.org/10.1016/j.ijsrc.2017.09.005>
- Salmasi, F., Shadkani, S., Abraham, J., Malekzadeh, F., 2022. Experimental Investigation for Determination of Discharge Coefficients for Inclined Slide Gates and Comparison with Data-Driven Models. *Iran. J. Sci. Technol. Trans. Civ. Eng.* 46, 2495–2509. <https://doi.org/10.1007/s40996-022-00850-9>
- Siasar, H., Honar, T., 2019. Application of Support vector machine, CHAID and Random forest models, in estimated daily Reference evapotranspiration in northern Sistan and Baluchestan province. *Iran. J. Irrig. Drain.* 13, 378–388.
- Taylor, K.E., 2001. Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res. Atmospheres* 106, 7183–7192. <https://doi.org/10.1029/2000JD900719>
- Valipour, M., 2013. INCREASING IRRIGATION EFFICIENCY BY MANAGEMENT STRATEGIES: CUTBACK AND SURGE IRRIGATION 8.



- Yan, X., Wang, Y., Fan, B., Mohammadian, A., Liu, J., Zhu, Z., 2023. Data-driven modeling of sluice gate flows using a convolutional neural network. *J. Hydroinformatics* 25, 1629–1647. <https://doi.org/10.2166/hydro.2023.200>
- Yang, L., Peng, T., Zheng, L., Wang, X., 2017. Experimental study on characteristics of discharge coefficient of hydraulic flap gate. *工程科学与技术 Adv. Eng. Sci.* 49, 54–59.
- Yaseen, Z.M., 2021. An insight into machine learning models era in simulating soil, water bodies and adsorption heavy metals: Review, challenges and solutions. *Chemosphere* 277, 130126. <https://doi.org/10.1016/j.chemosphere.2021.130126>
- Yaseen, Z.M., Alawi, O.A., Alshammari, A.M., Alsuwaiyan, A., Oyedeji, M.O., Oudah, A.Y., 2023. Development of Advanced Data-Intelligence Models for Radial Gate Discharge Coefficient Prediction: Modeling Different Flow Scenarios. *Water Resour. Manag.* 37, 5677–5705. <https://doi.org/10.1007/s11269-023-03624-8>
- Zayeri, M., 2023. Discharge Prediction in Flumes with Trapezoidal Contraction by Machine Learning Techniques. *Irrig. Drain. Struct. Eng. Res.* 24, 55–70. <https://doi.org/10.22092/idser.2023.363054.1549>
- Zhang, Y., Chiew, F.H.S., Li, M., Post, D., 2018. Predicting Runoff Signatures Using Regression and Hydrological Modeling Approaches. *Water Resour. Res.* 54, 7859–7878. <https://doi.org/10.1029/2018WR023325>