

مدل‌سازی ضریب دبی سرریز کرامپ با استفاده از روش‌های یادگیری ماشینی

محمدتقی ستاری^{۱*}، فرزین سلماسی^۲، امیر حاجی محمدی^۳

تاریخ دریافت: ۹۴/۰۳/۲۲ تاریخ پذیرش: ۹۵/۰۵/۳۰

^۱ - استادیار گروه مهندسی آب، دانشکده کشاورزی، دانشگاه تبریز

^۲ - دانشیار گروه مهندسی آب، دانشکده کشاورزی، دانشگاه تبریز

^۳ - کارشناس ارشد عمران آب، دانشگاه آزاد اسلامی واحد مراغه

* مسئول مکاتبات، پست الکترونیکی: mtsattar@gmail.com

چکیده

سرریز کرامپ در طبقه‌بندی سرریزها در میان سرریزهای لبه‌کوتاه جای می‌گیرد. طراحی این سرریز به‌گونه‌ای است که شیب بالادست آن بیشتر از شیب پایین‌دست آن بوده و موجب می‌شود که رسوبات به راحتی تخلیه گردند. در این تحقیق عملکرد روش‌های k-نزدیک‌ترین همسایگی و رگرسیون بردار پشتیبان در مدل‌سازی ضریب دبی سرریز کرامپ با استفاده از داده‌های آزمایشگاهی مورد بررسی قرار گرفت. داده‌ها در ۱۷۴ دسته و ۹ ترکیب مختلف از پارامترهای ورودی شامل شیب بالادست (S_{up})، شیب پایین‌دست (S_{do})، عدد رینولدز (Re) و نسبت عمق آب روی سرریز در بالادست به ارتفاع سرریز (h_1/P) مورد استفاده قرار گرفت. در چهار مرحله به ترتیب ۶۶، ۷۰، ۷۵ و ۸۰ درصد داده‌های آزمایشگاهی برای آموزش و مابقی آن‌ها در هر مرحله جهت آزمون استفاده گردید. براساس یافته‌ها در هر دو روش، بهترین نتیجه زمانی حاصل می‌شود که از ۸۰ درصد داده‌ها برای آموزش و ۲۰ درصد داده‌ها برای آزمون استفاده شود. از سوئی دیگر روش نزدیک‌ترین همسایگی در مقایسه با رگرسیون بردار پشتیبان از توانمندی بیشتری در مدل‌سازی ضریب دبی سرریز کرامپ برخوردار بوده و نسبت (h_1/P) تأثیر مهمی در تخمین ضریب دبی سرریز کرامپ داشته و قادر است با دقت نسبتاً بالایی ضریب دبی سرریز را مدل‌سازی نماید. همچنین نتایج نشان داد با کاربرد ترکیب ورودی شامل پارامترهای h_1/P ، S_{up} ، S_{do} ، روش‌های رگرسیون بردار پشتیبان و نزدیک‌ترین همسایگی به ترتیب با ضرایب همبستگی ۰/۹۶۹ و ۰/۹۸۷ بیشترین دقت را از خود نشان دادند.

واژه‌های کلیدی: رگرسیون بردار پشتیبان، سرریز کرامپ، ضریب دبی، k-نزدیک‌ترین همسایگی

Modeling Crump Weir Discharge Coefficient Using Machine Learning Methods

MT Sattari^{*1}, F Salmasi², A Hajimohammadi³

Received: 12 June 2015 Accepted: 20 August 2016

¹ Assis. Prof., Dept. of Water Engin., Faculty of Agric., Univ. of Tabriz, Tabriz, Iran

² Assoc. Prof., Dept. of Water Engin., Faculty of Agric., Univ. of Tabriz, Tabriz, Iran

³ M.Sc. in Civil Engin.-Water, Islamic Azad Univ., Maraghe Branch, Maraghe, Iran

*Corresponding Author, Email: mtsattar@gmail.com

Abstract

Crump weir is classified as a short-edged weir. The upstream slope of this weir is greater than its downstream slope, which allows easy discharging of the sediments. In this research, the performance of k-nearest neighborhood and support vector regression (SVR) methods were investigated for modelling crump weir discharge coefficient using experimental data. 174 data sets in 9 combinations of the input parameters including the upstream and downstream slopes (S_{up} , S_{do}), Reynolds number (R_e) and water head in upstream to weir height ratio (h_1/P) were used for modelling the discharge coefficient. The training was done in four stages using 66, 70, 75 and 80 percent of experimental data and the rest of these data at each stage were applied for the test phase. According to the results, the highest accuracy for the both applied models was obtained using 80% of the data in the training and the rest 20% in the test phases. Also, this investigation showed that the nearest neighborhood method presented a more accurate result than SVR method. Furthermore, water head in upstream to weir head height ratio (h_1/P) had a significant role in modeling crump weir discharge coefficient. This ratio was the only parameter which could be used for predicting the coefficient accurately. Finally, this work showed that input combination including h_1/P , S_{up} , S_{do} parameters gave the best outcome. Both the Nearest Neighborhood and Support Vector Regression methods with coefficient of determination values of 0.987 and 0.969 respectively, provided accurate predictions.

Keywords: Crump weir, Discharge coefficient, K-nearest neighborhood, Support vector regression

مقدمه

کردن آن موجب افزایش ضریب دبی جریان می‌شود (ترکمان ۱۳۹۲). یکی از دلایل استفاده از سرریز لبه‌پهن با وجوه شیب‌دار بالادست و پایین‌دست کاهش جدایی جریان است. بدین ترتیب که آزمایش‌ها نشان داده وجه عمودی در سرریز لبه‌پهن به دلیل ایجاد جدایی جریان به‌خصوص در دبی‌های زیاد، موجب کاهش ۱۰ تا ۱۵ درصدی ضریب دبی جریان می‌گردد (سارگیسون و پرسی ۲۰۰۹). به دلیل این‌که در سرریزها دو فاز آب‌وهوا در اختلاط هستند، این امر سبب می‌شود که وضعیت جریان پیچیده باشد و به همین دلیل، برای تحلیل این‌گونه جریان‌ها از مدل‌های فیزیکی استفاده می‌شود. از طرفی

در راستای بهره‌برداری بهینه از منابع آب به‌ویژه در مناطقی که با کمبود آب مواجه هستند، اندازه‌گیری دقیق آب عبوری از سازه‌های آبی از اهمیت شایانی برخوردار است. سرریزها از رایج‌ترین ابزار اندازه‌گیری دبی جریان در سازه‌های آبی هستند. از این نظر بررسی و تحقیق در مورد آن‌ها یکی از مسائل مهم در مهندسی آب می‌باشد. یکی از رایج‌ترین این سرریزها، سرریز کرامپ می‌باشد که در مقطع طولی دارای شکل مثلثی و در پلان دارای شکل مستطیلی است. شیب‌دار کردن وجه بالادست یا پایین‌دست سرریز لبه‌پهن و نیز گردگوشه

دوبعدی شبیه‌سازی شد و سطح آزاد جریان با روش جزء حجم سیال تعیین گردید. نتایج مدل عددی برای پارامترهای هیدرولیکی ضریب دبی، سرعت جریان و توسعه لایه‌مرزی با نتایج آزمایشگاهی و روابط موجود مقایسه شد. نتایج به‌دست‌آمده نشان داد که مدل آشفتگی $k-\epsilon$ و روش جزء حجم سیال VOF برای پیش‌بینی پروفیل سطح آب روی سرریزهای لبه‌پهن، توزیع سرعت و برآورد ضریب دبی مناسب می‌باشند. بیشینه خطای نسبی برای بار هیدرولیکی و ضریب دبی به‌ترتیب برابر با $4/06$ و $6/54$ درصد است که قابل‌قبول می‌باشد. سارگیسون و پرسی (۲۰۰۹) برای شبیه‌سازی جریان روی سرریز لبه‌پهن با جوه شیب‌دار از مدل‌های فیزیکی استفاده نمودند. آزمایش‌ها بر روی فلومی با کف افقی به‌عرض $0/2$ متر، ارتفاع $0/4$ متر و طول $5/4$ متر انجام‌شده است. سطح آب در پایاب در سطح پایین نگاه‌داشته شد تا بر روی جریان سرریز تأثیر نگذارد. سرریز از سه قسمت وجه بالادست، وجه پایین‌دست و تاج مستطیلی شکل تشکیل‌شده بود. وجه پایین‌دست و بالادست مدل‌ها با ترکیب شیب‌های مختلف $1:1$ و $2:1$ (عمودی: افقی) و عمودی ساخته شد. نتایج نشان داد که افزایش شیب بالادست به‌سمت حالت عمودی، موجب کاهش ارتفاع پروفیل سطح آب و نیز فشار استاتیک روی تاج گردید. این وضعیت همچنین باعث کاهش ضریب دبی شد. تغییرات شیب پایین‌دست اثر ناچیزی روی پروفیل سطح آب و فشار روی سرریز داشت. احتمال وقوع خلأزایی در کنار شیب پایین‌دست در حالت دبی‌های زیاد و شیب‌دار کردن جوه وجود دارد. شبیه‌سازی عددی جریان روی سرریزهای لبه‌پهن دوزنقه‌ای شکل توسط هان و همکاران (۲۰۱۱) انجام گرفت. در این تحقیق از دو مدل عددی Flow-3D و SSIM2 برای بررسی پروفیل سطح آب استفاده شد. اختلاف بین مقادیر محاسبات عددی با مقادیر مشاهداتی حاصل از مدل‌های فیزیکی برابر 1 تا $1/5$ درصد به‌دست آمد. دقت روش‌های عددی

ساخت مدل‌های فیزیکی نیز نیاز به‌صرف هزینه‌های بالایی دارد. این مدل‌ها از خطا مصون نبوده و ممکن است در معرض خطاهای انسانی و آزمایشگاهی قرار گیرند. اما امروزه با گسترش فناوری‌های رایانه‌ای و روش‌های محاسباتی و همچنین دینامیک سیالات محاسباتی این کار آسان و کم‌هزینه شده است. ازجمله این روش‌ها، می‌توان به روش‌های داده‌محور مبتنی بر یادگیری ماشینی اشاره نمود که در آن از داده‌های تجربی حاصل از اندازه‌گیری استفاده می‌گردد. تاکنون مطالعات گسترده‌ای در زمینه مدل‌سازی ضریب دبی در سرریزها صورت گرفته است که به برخی از آن‌ها به‌اختصار اشاره می‌شود. حبیبی و خسروجردی (۱۳۸۷) در مطالعه‌ای عبور جریان از روی سرریز کرامپ را به‌وسیله نرم‌افزار Fluent شبیه‌سازی نموده و با سرریز اوجی سد کوچری مقایسه نمودند. نتیجه حاصل حاکی از این بود که در سرریز کرامپ با افزایش ارتفاع آب روی سرریز، ضریب آبگذری کاهش می‌یابد. حاجی‌پور (۱۳۸۸) به بررسی مشخصه‌های جریان در سرریز کرامپ با ساخت ۸ مدل فیزیکی پرداخت. شکری و همکاران (۱۳۸۷) تأثیر شیب‌دار کردن وجه بالادست و پایین‌دست سرریزهای لبه‌پهن مستطیلی بر ضریب دبی و مشخصات جریان را موردبررسی قرار دادند. بررسی‌ها نشان داد که ضریب دبی جریان سرریز با تغییرات شیب پایین‌دست تقریباً ثابت می‌ماند. درحالی‌که با تغییر شیب وجه بالادست سرریز، ضریب دبی جریان و در نتیجه ظرفیت تخلیه سرریز متناسب با کاهش شیب بالادست افزایش می‌یابد. ترکمان (۱۳۹۲) با استفاده از داده‌های حاجی‌پور (۱۳۸۸) و بر اساس مدل درختی M5 اقدام به مدل‌سازی ضریب دبی در سرریز کرامپ نمود. وی نتیجه گرفت که مدل درختی M5 می‌تواند با ارائه روابط ساده خطی و با دقت بالایی ضریب دبی در سرریز کرامپ را محاسبه نماید. در تحقیق عباسپور و همکاران (۱۳۹۲) جریان روی دو نوع سرریز لبه‌پهن BRA و ARB با وجه‌های شیب‌دار پایین‌دست و بالادست و به‌کارگیری نرم‌افزار فلوئنت به‌صورت

۱:۵)، (۲:۳ و ۲:۹) و ارتفاع های ۱۷، ۱۶، ۱۵ سانتی متر صورت گرفته است. در شکل ۱ مقطع طولی از سرریز کرامپ و در جدول ۱ ویژگی های آماری پارامترهای ورودی و خروجی مورد استفاده در این تحقیق ارائه شده است.

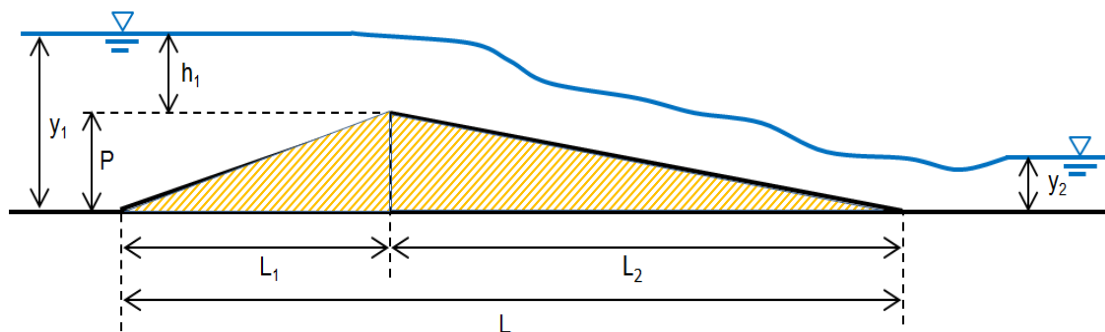
در این تحقیق داده های آزمایشگاهی در ۱۷۴ دسته و ۹ ترکیب مختلف از پارامترهای ورودی (شیب بالادست، شیب پایین دست، عدد رینولدز و نسبت عمق آب روی سرریز در بالادست به ارتفاع سرریز) جهت مدل سازی مورد استفاده قرار گرفت. دلیل انتخاب ۹ ترکیب از پارامترهای ورودی جهت مدل سازی، مقادیر همبستگی (شکل ۳) بین پارامترهای ورودی و ضریب دبی سرریز به عنوان پارامتر خروجی و یا ویژگی هدف در نظر گرفته شد. در مدل های داده محور اگر چه محدودیت صریحی برای تعداد داده ها در فرآیند مدل سازی وجود ندارد، ولیکن طبیعی است که زیاد بودن سری داده ها می تواند موجب بهبود نتایج مدل سازی گردد. در این تحقیق محدودیت های آزمایشگاهی باعث گردید تا صرفاً از ۱۷۴ دسته داده موجود جهت مدل سازی استفاده گردد. شکل ۲ فراوانی داده های مورد استفاده برای هر یک از پارامترهای مورد استفاده را نشان می دهد. چنان که از شکل ۲ به صورت بصری استنباط می شود، توزیع داده های مربوط به ضریب دبی سرریز به عنوان ویژگی هدف در فرآیند مدل سازی، نسبت به بقیه پارامترها به توزیع نرمال نزدیک تر بوده و ایده استواری نتایج مستخرج از مدل ها را تقویت می کند.

بستگی به اندازه شبکه ها داشت. تعداد مناسب شبکه ها برای دو مدل به ترتیب ۶۳۳۳ و ۹۷۶۹ عدد به دست آمد. بر این اساس در تحقیق حاضر با مقایسه نتایج حاصل از دو روش k -نزدیک ترین همسایگی و رگرسیون بردار پشتیبان، میزان دقت آن ها در تعیین ضریب دبی سرریز کرامپ مورد بررسی قرار گرفته و تأثیر هر یک از پارامترهای مستقل ورودی مدل، در تخمین ضریب دبی سرریز کرامپ ارزیابی می گردد. داده های مربوط به ضریب دبی سرریز کرامپ، مورد استفاده در این تحقیق از آزمایش ها روی مدل های فیزیکی حاجی پور (۱۳۸۸) اخذ گردیده اند.

مواد و روش ها

داده های آزمایشگاهی مورد استفاده

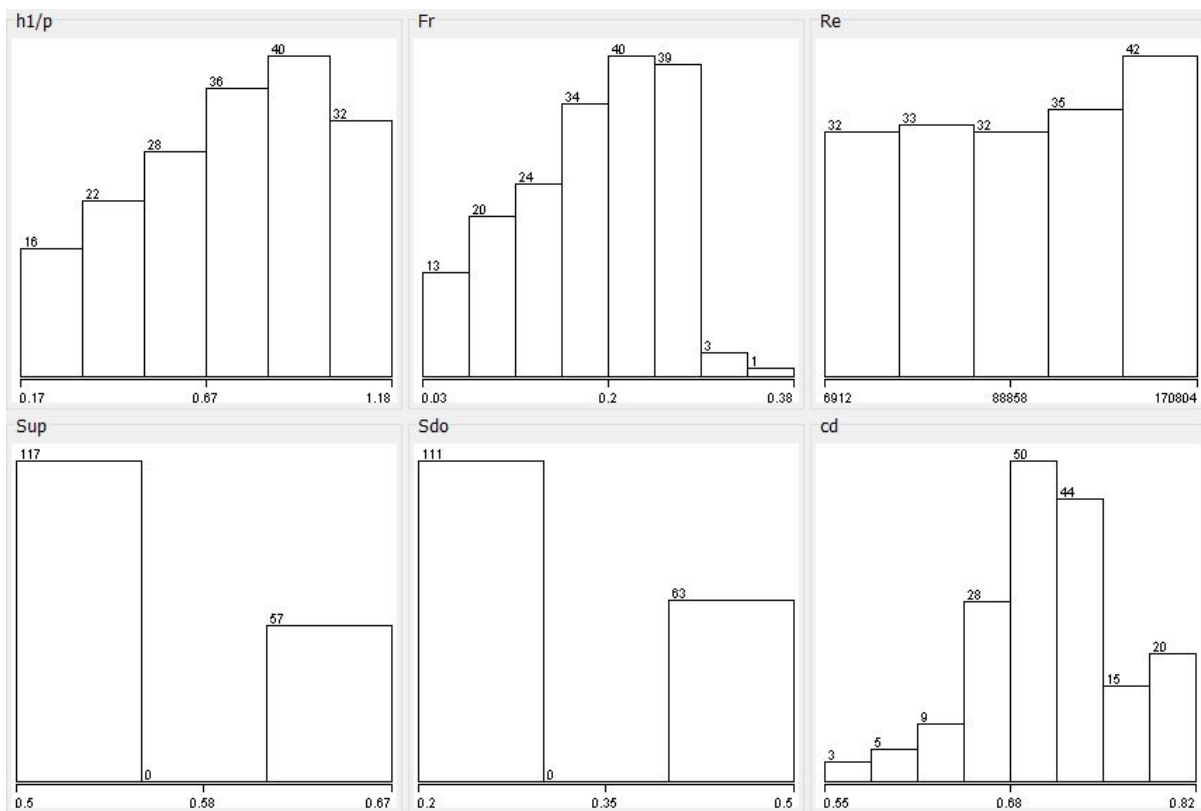
داده های آزمایشگاهی مورد استفاده در این تحقیق برگرفته از ۹ مدل فیزیکی مربوط به سرریز کرامپ می باشد (حاجی پور ۱۳۸۸). این آزمایش ها در آزمایشگاه هیدرولیک دانشکده کشاورزی دانشگاه تبریز انجام گرفته است که با تغییر ابعاد سازه های سرریز (شیب بالادست و پایین دست سرریز)، ارتفاع سرریز به ازای دبی در واحد عرض های مختلف (در بازه $0/172 - 0/0696$ مترمربع بر ثانیه)، پارامترهای هیدرولیکی جریان در نقاط مختلف روی سرریز اندازه گیری شده است. آزمایش در کانالی به طول ۱۰ متر، عرض ۲۵ سانتی متر و ارتفاع ۵۰ سانتی متر و سرریزی با شیب های بالادست و پایین دست (۱:۲ و ۱:۲)، (۱:۲ و ۱:۲) و



شکل ۱- مقطع طولی از سرریز کرامپ و معرفی پارامترهای هیدرولیکی آن.

جدول ۱- ویژگی آماری پارامترهای ورودی و خروجی مورداستفاده در این تحقیق.

h ₁ /p (عمق آب روی سرریز در بالادست به ارتفاع سرریز)	Re (عدد رینولدز)	S _{up} (شیب بالادست)	S _{do} (شیب پایین دست)	C _d (ضریب دبی)	آماره
۰/۱۶۵	۶۹۱۲	۰/۵	۰/۲	۰/۵۴۹	کمینه
۱/۱۸	۱۷۰۸۰۴	۰/۶۶۷	۰/۵	۰/۸۱۶	بیشینه
۰/۷۴۲	۹۳۳۴۷	۰/۵۵۵	۰/۳۱۶	۰/۷۱۱	میانگین
۰/۲۶۸	۴۸۰۷۹	۰/۰۷۹	۰/۱۳۹	۰/۰۵۱	انحراف معیار



شکل ۲- فراوانی داده‌های مورداستفاده در این تحقیق.

و شیب‌های بالادست و پایین دست از کمترین همبستگی نسبت به ضریب دبی برخوردار هستند. علت آن را می‌توان این‌گونه بیان نمود که اصولاً در سازه‌های هیدرولیکی، عوامل هیدرولیکی جریان مانند دبی، سرعت جریان و عمق آب جزء عوامل تأثیرگذارتر نسبت به عوامل هندسی سازه مانند شیب و ارتفاع سازه هستند.

در مورد سرریز کرامپ نیز این وضعیت صادق بوده است. ناچیز بودن تغییرات شیب روی ضریب دبی، قبلاً توسط سارگیسون و پرس (۲۰۰۹) نیز اعلام شده

جهت انجام مدل‌سازی و تعیین ترکیب پارامترهای ورودی هر یک از سناریوها نیاز به سنجش تأثیر و وابستگی هر یک از پارامترهای مستقل بر روی پارامتر هدف یعنی ضریب دبی سرریز کرامپ وجود دارد. شکل ۳ همبستگی بین پارامترها را به صورت بصری نشان می‌دهد.

همچنان که از شکل ۳ ملاحظه می‌شود عدد رینولدز و همچنین پارامتر نسبت عمق آب روی سرریز در بالادست به ارتفاع سرریز (h₁/P) بیشترین همبستگی

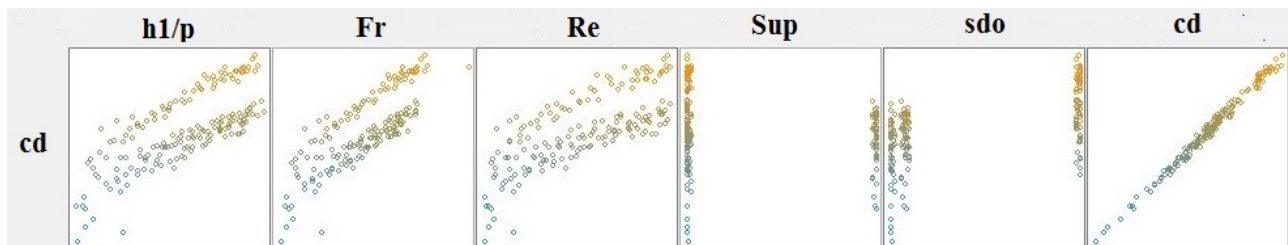
نمود که داده‌ها در یک فضای دوبعدی قرار دارند و بنابراین بایستی به دنبال خطی جهت جداسازی کلاس‌ها بود. هنگامی که فضا سه‌بُعدی می‌شود، جداکننده یک صفحه است و اگر تعداد صفات خاصه در مجموعه داده‌های اولیه بیش از سه عدد باشد، از کلمه ابر صفحه برای جداسازی میان کلاس‌ها استفاده می‌شود. در طبقه‌بندی خطی، مجموعه صفات خاصه توسط بردار X_i و برجسب کلاس با متغیر Y_i نمایش داده می‌شود و مقادیر $+1$ و -1 دو برجسب کلاس را تشکیل می‌دهند که داده‌های هر کلاس حق هیچ‌گونه تخطی به کلاس دیگر را ندارند. با فرض اینکه ورودی داده‌ها دوبعدی است، بایستی به دنبال ابر صفحه‌ای بود که بتواند دو کلاس موردنظر را به نحوی تفکیک نماید که فاصله بین ابر صفحه و نزدیک‌ترین نمونه آموزشی به حالت بهینه برسد. منظور از رگرسیون، به دست آوردن یک ابر صفحه است که بر داده‌های موردنظر برازش داده می‌شود. فاصله هر نقطه از این ابر صفحه نشان‌دهنده خطای آن نقطه خاص است. حل معادله پیدا کردن ابر صفحه بهینه برای داده‌ها به وسیله روش‌های برنامه‌ریزی درجه دوم (QP) و ضرایب لاگرانژ که روش‌های شناخته شده‌ای در حل مسائل محدودیت‌دار هستند، صورت می‌گیرد. مهم‌ترین پارامتر مؤثر در رگرسیون بردار پشتیبان، پیدا نمودن مقدار بهینه برای ضرایب لاگرانژ است.

است. لازم به ذکر است که تعداد شیب‌های مورد استفاده در این تحقیق به خاطر محدودیت‌های فیزیکی و آزمایشگاهی مقادیر (۱:۲ و ۱:۲)، (۱:۲ و ۱:۵)، (۲:۳ و ۲:۹) در نظر گرفته شد.

جهت انجام مدل‌سازی و آزمون تأثیر همه پارامترهای مؤثر در مدل‌سازی ضریب دبی سرریز کرامپ، ترکیبات مختلفی از پارامترهای ورودی در نظر گرفته شد. در همین راستا ابتدا داده‌ها در هر ترکیب به دو گروه آموزش و آزمون تقسیم شد. این تقسیم‌بندی در چهار مرحله، به صورت ۸۰-۷۵-۷۰-۶۶ درصد داده‌ها جهت آموزش و ۲۰-۲۵-۳۰-۳۴ درصد داده‌ها جهت آزمون مدل مدنظر قرار گرفت. جهت پرهیز از اطاله کلام صرفاً نتایج بهترین تقسیم‌بندی که ۸۰-۲۰ بود، در این مقاله ارائه گردید.

رگرسیون بردار پشتیبان

ماشین بردار پشتیبان یکی از روش‌های یادگیری با نظارت است که از آن برای طبقه‌بندی و رگرسیون استفاده می‌کنند. مبنای کاری دسته‌بندی‌کننده ماشین بردار پشتیبان دسته‌بندی خطی داده‌ها است و در تقسیم خطی داده‌ها سعی می‌شود خطی انتخاب شود که حاشیه اطمینان بیشتری داشته باشد (کریستوفر و بورگز ۱۹۹۸). در این روش، داده‌ها به دودسته آموزش و آزمون تقسیم می‌شوند. چنانچه مجموعه داده‌های آموزشی دارای دو صفت خاصه باشند، می‌توان تصور



شکل ۳- ماتریس همبستگی بصری بین پارامترهای ورودی و خروجی.

الگوریتم ساده‌ای است که می‌تواند مسئله QP را به سرعت بدون ذخیره‌سازی ماتریسی اضافی و هیچ‌گونه استفاده‌ای از مراحل بهینه‌سازی عددی QP حل کند. دو

بنابراین با توجه به اهمیت موضوع، در این تحقیق از روش بهینه‌سازی کمینه‌ساز ترتیبی (SMO) استفاده شده است. بهینه‌سازی کمینه‌ساز ترتیبی،

بر اساس فاصله و انتخاب k همسایه نزدیک‌تر، ج) استفاده از دسته‌ای که قسمت عمده را در همسایه‌های نزدیک، به‌عنوان تخمینی برای دسته نمونه دارد، تشکیل می‌شود. برای محاسبه فاصله میان دو نمونه از توابع مختلفی استفاده می‌شود که در این تحقیق از فاصله اقلیدسی استفاده شده است. نحوه محاسبه فاصله بر اساس روش‌های مختلف در روابط ۱ تا ۳ ارائه گردیده است. در این روابط x_1 و x_2 بیان‌گر نقاط و D فاصله بین آن‌ها می‌باشد.

بعد از مشخص شدن نزدیک‌ترین همسایه‌ها، الگوریتم باید از روی دسته آن‌ها، دسته نمونه ورودی را تعیین کند که به این عمل ترکیب گویند. ساده‌ترین روش ترکیب، روش بدون وزن است که بر مبنای رأی قسمت عمده است، یعنی کلاسی که حائز قسمت عمده در بین نزدیک‌ترین همسایه‌ها باشد انتخاب می‌شود. یکی از پارامترهای مهم در روش نزدیک‌ترین همسایگی، مقدار k می‌باشد. واقعیت این است که مقدار دقیقی برای k وجود نداشته و مقدار مناسب آن بستگی به توزیع داده‌ها و فضای مسئله دارد. ولی مقدار کوچک k ، روش را متأثر از داده‌های مغشوش کرده و مقدار بزرگ آن، رفتارهای محلی را در نظر نمی‌گیرد. نهایتاً مقدار k ، با سعی و خطا به دست می‌آید. مثلاً در روش اعتبار سنجی متقاطع، با مقدار اولیه شروع کرده و k تغییر داده می‌شود تا خطای دسته‌بندی به کمترین مقدار ممکن برسد. اگرچه روش نزدیک‌ترین همسایگی در مطالعات هیدرولوژیکی (عزمی و عراقی‌نژاد ۱۳۹۲) بیشتر استفاده گردیده است ولیکن محدودیتی در استفاده از این روش برای مطالعات آزمایشگاهی و هیدرولیکی وجود نداشته و مطالعاتی در این زمینه گزارش گردیده است (فدائی کرمانی و همکاران ۱۳۹۱). در این تحقیق از نرم‌افزار WEKA برای مدل‌سازی با روش نزدیک‌ترین همسایگی استفاده گردیده است.

$$D(x_1, x_2) = \max_i (|x_{1i} - x_{2i}|) \quad [1]$$

فاصله
چیبی‌چف

مؤلفه در SMO وجود دارد: ۱) روش تحلیلی برای حل دو ضریب لاگرانژی و ۲) روش ابتکاری برای انتخاب ضرایب به‌منظور بهینه‌سازی (شهرابی و ذوقدر شجاعی ۱۳۹۰). قبل از تقسیم خطی برای اینکه ماشین بتواند داده‌های با پیچیدگی بالا را دسته‌بندی کند، داده‌ها به فضای با ابعاد خیلی بالاتر برده می‌شود. در مسائل واقعی منطقی نیست که فرض کنیم تمامی نقاط به‌صورت خطی تفکیک‌پذیر هستند. دسته‌بندی غیرخطی پیچیده یکی از این مسائل واقعی می‌باشد. برای اینکه بتوان مسئله ابعاد خیلی بالا را حل نمود، از قضیه دوگانی لاگرانژ و از توابع هسته مختلفی از جمله هسته‌های نمایی، چندجمله‌ای و سیگموئید می‌توان استفاده نمود (کریستوفر و بورگز ۱۹۹۸). توابع هسته (کرنل) براساس ضرب داخلی داده‌های مفروض، تعریف می‌شود. درواقع، توابع کرنل با یک تبدیل غیرخطی از فضای ورودی به فضای خصیصه با ابعاد بیشتر (حتی نامتناهی) می‌توانند مسائل را به‌صورت خطی تفکیک‌پذیر نمایند. در این تحقیق پس از آزمون توابع کرنل مختلف از تابع پلی‌کرنل جهت خطی‌سازی استفاده گردید. کلیه مراحل مدل‌سازی رگرسیون بردار پشتیبان بر اساس روش بهینه‌سازی کمینه‌ساز ترتیبی (SMO) با استفاده از نرم‌افزار WEKA (ویتن و همکاران ۲۰۱۱) انجام پذیرفت.

روش K- نزدیک‌ترین همسایگی (K-NN)

روش KNN از ابتدای سال ۱۹۷۰ به‌عنوان یک روش غیر پارامتری مورد استفاده قرار گرفته است، که هدف کلی آن طبقه‌بندی و تخمین ویژگی‌های یک سری داده‌های مجهول با توجه به بیشترین شباهت این داده‌ها با داده‌های معلوم که در همسایگی (نزدیکی) آن‌ها قرار دارند، می‌باشد (براندسما و اونن ۲۰۰۶). این الگوریتم تا ورود داده‌ی آزمایشی صبر می‌کند و به ساخت مدلی جهت طبقه‌بندی نمی‌پردازد، به‌همین دلیل به یادگیرنده‌های تنبل موسوم است. الگوریتم نزدیک‌ترین همسایگی از سه‌گام الف) محاسبه فاصله نمونه ورودی با تمام نمونه‌های آموزشی، ب) مرتب کردن نمونه‌های آموزشی

نتایج و بحث

در این تحقیق پس از انجام پیش‌پردازش و آماده‌سازی داده‌ها، عملکرد روش‌های k-نزدیک‌ترین همسایگی و رگرسیون بردار پشتیبان در مدل‌سازی ضریب دبی سرریز کرامپ با استفاده از داده‌های آزمایشگاهی مورد بررسی قرار گرفت و نتایج به‌دست‌آمده از فرآیند مدل‌سازی تحلیل و به‌صورت جداول و نمودارهای مربوطه ارائه گردید. در مدل‌سازی با روش رگرسیون بردار پشتیبان در تابع هدف کمینه‌سازی مربوط به ابر صفحه بهینه در صورتی که داده‌ها از مقدار مجاز تخطی نمایند می‌بایست متناسب با مقدار تخطی، متغیر کمبود تعریف شوند (شهرابی و نولقدر شجاعی ۱۳۹۰). این متغیرهای کمبود، متغیرهای کمکی هستند و پارامتر C ضریب تعادل پیچیدگی میان ماشین و تعداد نقاط تفکیک‌ناپذیر است که با سعی و خطا به‌دست می‌آید. در تحقیق حاضر مقدار C پس از آزمون و خطا برابر ۰/۷ انتخاب شد. مقدار تابع زیان ϵ -غیرحساس (Insensitive loss function) نیز در نرم‌افزار WEKA به‌طور پیش‌فرض برابر ۰/۰۰۱ می‌باشد که پس از آزمون‌های انجام‌گرفته مقدار پیش‌فرض به‌عنوان عدد بهینه انتخاب شد. در فرآیند مدل‌سازی با روش نزدیک‌ترین همسایگی تعداد همسایه‌های مختلف مورد آزمون قرار گرفت و مشخص گردید که بهترین نتایج در بازه تعداد همسایه‌های ۳ تا ۶ به‌دست می‌آید. جدول ۲ نتایج به‌دست‌آمده برای تمامی ترکیب پارامترهای ورودی را با استفاده از هر دو روش مدل‌سازی نشان می‌دهد. لازم به ذکر است که در جدول ۲ منظور از K همان تعداد همسایه می‌باشد.

$$D(x_1, x_2) = \sqrt{\sum_{i=1}^n (x_{1i} - x_{2i})^2} \quad \begin{array}{l} \text{فاصله} \\ \text{اقلیدسی} \end{array} \quad [2]$$

$$D(x_1, x_2) = \sum_{i=1}^n |x_{1i} - x_{2i}| \quad \begin{array}{l} \text{فاصله مَنهتن} \end{array} \quad [3]$$

معیارهای ارزیابی مدل

در این تحقیق جهت سنجش عملکرد مدل‌ها از آماره‌های ضریب همبستگی، میانگین قدر مطلق خطا و جذر میانگین مربعات خطا استفاده شده است. ضریب همبستگی، نوع و میزان همبستگی موجود بین داده‌های ورودی و خروجی مدل را تعیین می‌کند و مقدار آن از رابطه ۴ محاسبه می‌شود:

$$R = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}} \quad [4]$$

که در آن x_i مقدار ورودی و y_i مقدار خروجی مدل، \bar{x} ، \bar{y} به‌ترتیب معرف مقدار میانگین داده‌های ورودی و خروجی است. میانگین قدر مطلق خطا و جذر میانگین مربعات خطا نشان‌دهنده میزان خطای مدل می‌باشند که بهترین مقدار آن‌ها برابر صفر بوده و از طریق روابط ۵ و ۶ به‌صورت زیر محاسبه می‌شوند. در این روابط N بیان‌گر تعداد نمونه‌هاست.

$$MAE = \frac{\sum_{i=1}^N |x_i - y_i|}{N} \quad [5]$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (x_i - y_i)^2}{N}} \quad [6]$$

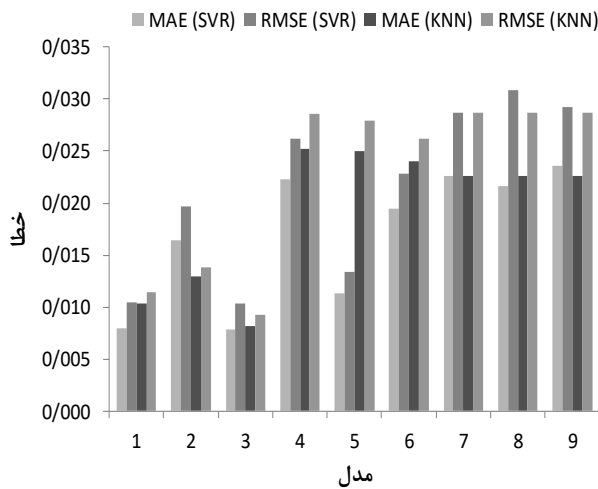
جدول ۲- نتایج مدل‌سازی با دو روش نزدیک‌ترین همسایگی و رگرسیون بردار پشتیبان برای ترکیب‌های مختلف ورودی.

مدل	ترکیب پارامترهای ورودی	SVR			KNN		
		R	MAE	RMSE	R	MAE	RMSE

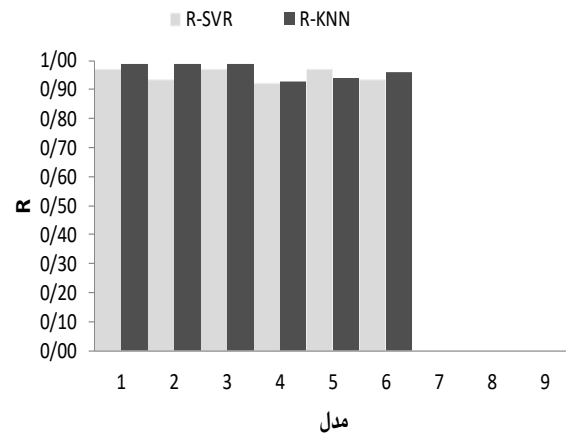
۱	$h_1/P, R_e, S_{up}, S_{do}$	۰/۹۶۹	۰/۰۰۸	۰/۰۱۱	۰/۹۸۶	۰/۰۱۰	۰/۰۱۱	۳
۲	R_e, S_{up}, S_{do}	۰/۹۳۲	۰/۰۱۶	۰/۰۲۰	۰/۹۸۷	۰/۰۱۳	۰/۰۱۴	۴
۳	$h_1/P, S_{up}, S_{do}$	۰/۹۶۹	۰/۰۰۸	۰/۰۱۰	۰/۹۸۷	۰/۰۰۸	۰/۰۰۹	۳
۴	$h_1/P, R_e$	۰/۹۳۲	۰/۰۲۲	۰/۰۲۶	۰/۹۳۸	۰/۰۲۵	۰/۰۲۹	۴
۵	h_1/P	۰/۹۶۹	۰/۰۱۱	۰/۰۱۳	۰/۹۳۹	۰/۰۲۵	۰/۰۲۸	۵
۶	R_e	۰/۹۳۲	۰/۰۲۰	۰/۰۲۳	۰/۹۶۱	۰/۰۲۴	۰/۰۲۶	۶
۷	S_{up}, S_{do}	۰/۰۰۰	۰/۰۲۳	۰/۰۲۹	۰/۰۰۰	۰/۰۲۳	۰/۰۲۹	-
۸	S_{up}	۰/۰۰۰	۰/۰۲۲	۰/۰۳۱	۰/۰۰۰	۰/۰۲۳	۰/۰۲۹	-
۹	S_{do}	۰/۰۰۰	۰/۰۲۴	۰/۰۲۹	۰/۰۰۰	۰/۰۲۳	۰/۰۲۹	-

به نزدیکی شرایط هیدرولیکی و آزمایشگاهی در حین انجام آزمایشات پی‌درپی توجه نمود. ترکیب‌های ورودی ۷ تا ۹ مربوط به سناریوهایی است که در آن تنها از پارامتر شیب برای مدل‌سازی استفاده شده است. در این سناریوها ضریب همبستگی برای هر دو روش صفر گردید و از این رو در ستون مربوط به تعداد همسایه‌ها هیچ همسایه‌ای در نظر گرفته نشد. سناریوهای ۷ تا ۹ نشان می‌دهد که شیب سرریز به‌تنهایی هیچ تأثیری در دقت مدل‌سازی ندارد. شکل‌های ۴ و ۵ تغییرات مربوط به ضرایب همبستگی و خطا به‌ازاء ترکیب‌های مختلف ورودی را نشان می‌دهند.

جدول ۲ نشان می‌دهد در مدل‌سازی با رگرسیون بردار پشتیبان (SVR) و روش نزدیک‌ترین همسایگی (KNN) سناریوی ۳ (شامل شیب‌های بالادست و پایین‌دست و نسبت عمق آب روی سرریز در بالادست به ارتفاع سرریز) به‌ترتیب با ضریب همبستگی ۰/۹۶۹ و ۰/۹۸۷ از دقت بیشتری برخوردار است. در سناریوی ۳ و در روش همسایگی به‌ازاء ۳ همسایه بیشترین دقت حاصل گردید. چنان‌که ملاحظه می‌شود روش نزدیک‌ترین همسایگی برتری نامحسوسی نسبت به روش رگرسیون بردار پشتیبان دارد. با توجه به اینکه روش نزدیک‌ترین همسایگی تنها با ۳ همسایه به جواب بهینه دست‌یافته است، شاید بتوان این برتری را با توجه



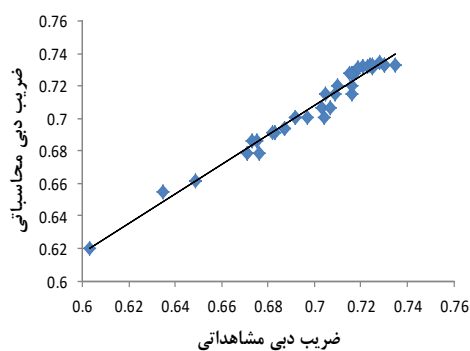
شکل ۵- نمودار تغییرات خطا.



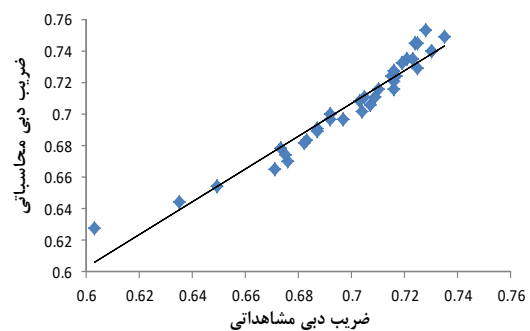
شکل ۴- نمودار تغییرات ضریب همبستگی.

سایر پارامترها قادر به بهبود نتایج مدل‌سازی می‌باشد. چنان‌که از شکل ۵ نیز پیداست روش نزدیک‌ترین همسایگی نسبت به روش رگرسیون بردار پشتیبان از کمترین خطا برخوردار است. نمودار پراکنش ضریب دبی محاسباتی برای هر دو روش مورد استفاده در این تحقیق در شکل‌های ۶ و ۷ ارائه شده است.

چنان‌که در شکل ۴ ملاحظه می‌شود در ترکیب-هایی که در آن شیب بالادست و یا پایین‌دست به‌تنهایی و یا توأم به‌عنوان ورودی مدل در نظر گرفته شده‌اند، نتیجه مطلوب حاصل نشده است. طبیعتاً به لحاظ هیدرولیکی واضح است که عامل شیب به‌تنهایی نمی‌تواند تأثیری در ضریب دبی داشته باشد ولیکن در ترکیب با



شکل ۷- نمودار پراکنش ضریب دبی برای ترکیب ورودی ۳ روش نزدیک‌ترین همسایگی.



شکل ۶- نمودار پراکنش ضریب دبی برای ترکیب ورودی ۳ روش رگرسیون بردار پشتیبان.

نسبت به روش رگرسیون بردار پشتیبان از دقت نسبتاً بالایی برخوردار است. با توجه به اینکه داده‌های تجربی مورد استفاده در این تحقیق بر اساس آزمایشات حاجی

چنان‌که از شکل‌های ۶ و ۷ استنباط می‌شود هر دو روش با دقت مناسبی می‌توانند ضریب دبی سرریز کرامپ را مدل‌سازی نمایند ولیکن روش نزدیک‌ترین همسایگی

ترکیب ورودی ۳ (شامل شیب‌های بالادست و پایین‌دست و نسبت عمق آب روی سرریز در بالادست به ارتفاع سرریز) به ترتیب با ضریب همبستگی ۰/۹۶۹ و ۰/۹۸۷ از دقت بیشتری برخوردار است. به نظر می‌رسد ترکیب ورودی ۵ از نظر کاربردی بودن نسبت به دیگر ترکیب‌های ورودی‌ها ارجحیت دارد. زیرا مشخص نمودن پارامتر نسبت عمق آب روی سرریز در بالادست به ارتفاع سرریز استفاده‌شده در این ترکیب ورودی، به لحاظ عملی ساده‌تر است. همچنین نتایج نشان‌داد به دلیل دامنه کم تغییرات شیب‌ها در حین انجام آزمایش‌ها مربوطه، شیب سرریز در بالادست و پایین‌دست در ترکیب با سایر پارامترها تأثیر بسیار اندکی روی ضریب دبی سرریز کرامپ داشته و می‌توان از اثر آن چشم‌پوشی کرد. در مطالعه‌ای که توسط ترکمان (۱۳۹۲) بر اساس داده‌های تجربی حاجی‌پور (۱۳۸۸) انجام پذیرفته، ضریب دبی سرریز کرامپ با ترکیب پارامترهای ورودی شامل Re ، Fr ، $d_{1/p}$ ، Sup (شیب بالادست) و S_{do} (شیب پایین‌دست) توسط مدل M5 محاسبه گردیده و ضریب همبستگی ۰/۹۸۸ به دست آمده، که بیانگر مشابهت جدی در نتایج هر دو تحقیق است.

پور (۱۳۸۸) می‌باشد، داده تجربی دیگری جهت مقایسه نتایج یافت نشد.

نتیجه‌گیری کلی

در این مقاله کاربرد روش رگرسیون بردار پشتیبان و k نزدیک‌ترین همسایگی در پیش‌بینی ضریب دبی سرریز کرامپ مورد بررسی قرار گرفت. برای نیل به هدف، از داده‌های تجربی حاجی‌پور (۱۳۸۸) استفاده گردید. نتایج به دست آمده نشان می‌دهد روش نزدیک‌ترین همسایگی در مقایسه با روش رگرسیون بردار پشتیبان از توانمندی بیشتری در مدل‌سازی ضریب دبی سرریز کرامپ برخوردار است. در این تحقیق گرچه از در صداهای متفاوتی برای تقسیم داده‌ها به دو بخش آموزش و آزمون استفاده شد ولی نتایج نشان داد زمانی که از ۸۰ درصد داده‌ها برای آموزش و ۲۰ درصد داده‌ها برای آزمون استفاده شود، بهترین نتیجه حاصل می‌شود و بیشترین تأثیر در مدل‌سازی ضریب دبی مربوط به پارامتر (نسبت عمق آب روی سرریز در بالادست به ارتفاع سرریز) می‌باشد. از سویی دیگر نتایج نشان می‌دهد در مدل‌سازی با رگرسیون بردار پشتیبان و روش نزدیک‌ترین همسایگی

منابع مورد استفاده

- شکری ن، فرهودی ج و گودرزی ا، ۱۳۸۷. تأثیر شیب وجه پایین‌دست و بالادست سرریز مستطیلی لبه‌پهن بر شرایط جریان. نشریه دانشکده فنی دانشگاه تبریز، دوره ۳۷. شماره ۳ (پیاپی ۵۵)، صفحه‌های ۹ تا ۱۹.
- شهرابی ج و ذوالقدر شجاعی ع، ۱۳۹۰. داده‌کاوی پیشرفته (مفاهیم و الگوریتم‌ها)، انتشارات جهاد دانشگاهی واحد صنعتی امیرکبیر، صفحه‌های ۴۵۷ تا ۴۹۲.
- عباسپور ا، عبدالله پور م و سلماسی ف، ۱۳۹۲. شبیه‌سازی عددی جریان روی سرریز لبه‌پهن مستطیلی با وجوه شیب‌دار بالادست و پایین‌دست با مدل فلونت. نشریه دانش آب‌و خاک، جلد ۲۳، شماره ۴، صفحه‌های ۲۶۵ تا ۲۷۶.
- عزمی م و عراقی‌نژاد ش، ۱۳۹۲. توسعه روش رگرسیون k -نزدیک‌ترین همسایگی در پیش‌بینی جریان رودخانه. آب و فاضلاب، دوره ۲۳، شماره ۲ (مسلسل ۸۲)، صفحه‌های ۱۰۷ تا ۱۱۸.
- فدائی کرمانی ا، بارانی غ و قائینی حصاروئی م، ۱۳۹۱. پیش‌بینی خطر وقوع جریان‌های گردابی در آبگیر قائم با استفاده از مدل نزدیک‌ترین همسایگی k . فصل‌نامه علمی پژوهشی مهندسی آبیاری و آب، سال ۳، شماره ۱۰، صفحه‌های ۹۲ تا ۱۰۰.

- ترکمان غ، ۱۳۹۲. ارزیابی عملکرد روش‌های ANFIS و مدل M5 در مدل‌سازی ضریب دبی در سرریز کرامپ. پایان‌نامه کارشناسی ارشد، دانشگاه آزاد اسلامی واحد اهر.
- حاجی‌پور ق ع، ۱۳۸۸. بررسی مشخصه‌های جریان در سرریز کرامپ با مدل فیزیکی و CFD. پایان‌نامه کارشناسی ارشد، گرایش سازه‌های آبی، دانشکده کشاورزی دانشگاه تبریز.
- حبیبی م و خسروجردی ا، ۱۳۸۷. بررسی ضریب آب‌گذری سرریز کرامپ و مقایسه آن با سرریز اوجی با استفاده از نرم‌افزار FLUENT. ۱۰ صفحه. سومین کنفرانس مدیریت منابع آب، ۲۳ تا ۲۵ مهرماه، دانشگاه تبریز، تبریز.
- Brandsma T and Onnen GPK, 2006. Application of nearest neighbor re-sampling for homogenizing temperature records on a daily to sub daily level. *International Journal of Climatology* 26(1): 75–89.
- Christopher J and Burges D, 1998. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery* 2(2): 121 – 167.
- Haun S, Reidar BN, Feurich O and Feurich R, 2011. Numerical modeling of flow over trapezoidal broad crested weir. *Engineering Application of Computational Fluid Mechanics* 5(3): 327-405.
- Sargison JE and Percy A, 2009. Hydraulics of broad-crested weirs with varying side slopes. *Journal of Irrigation and Drainage Engineering, ASCE* 135(1): 115-118.
- Witten IH, Eibe F and Mark H, 2011. *Data Mining: Practical Machine Learning Tools and Techniques*. San Francisco, Morgan Kaufmann 629p.