

هم‌چنین روش‌های داده‌کاوی نسبت به روش رگرسیونی دقت بالاتری دارند.

واژه‌های کلیدی: آبخیز، آب‌شکن، دیوار جداکننده، رسوب، داده‌کاوی.

## مقدمه

فرایند آبخیزداری از رودخانه‌ها به وسیله سازه آبخیز همواره با انحراف رسوبات با اندازه‌های متفاوت همراه می‌باشد. انحراف رسوبات به صورت بار بستر به کانال آبخیز، تأسیسات انتقال آب پایین دست آبخیز و طرح آب‌رسانی را با مشکلاتی روبرو می‌نماید. اصلاح ساختار جریان با استفاده از سازه‌های حفاظتی از روش‌های افزایش آبخیزداری و کاهش رسوب ورودی به آبخیز می‌باشد. سازه‌های مختلفی مانند صفحات مستغرق، آستانه، آب‌شکن، ترکیب آن‌ها و دیوار جداکننده برای کنترل رسوب در آبخیزها به کاررفته است. تاکنون تحقیقات زیادی روی صفحات مستغرق، آستانه، آب‌شکن و ترکیب آن‌ها با یکدیگر انجام شده است. به‌طورکلی پیچیدگی و سه‌بعدی بودن جریان در محدوده دهانه آبخیز، پیچیدگی پدیده حمل رسوب، تأثیر کاهش رسوبات ورودی به آبخیز و کاهش ظرفیت حمل رسوب در کانال اصلی بر مورفولوژی بستر، منجر به تحقیقات زیادی در این زمینه شده است. در این تحقیق تأثیر زاویه و ابعاد دیوار جداکننده و آب‌شکن بر روی کنترل رسوب و راندمان آبخیزداری به صورت آزمایشگاهی و استفاده از روش‌های نرم داده‌کاوی و رگرسیونی مورد بررسی قرار گرفته است. امروزه سیستم‌های هوشمند فازی و عصبی با توجه به توانایی در حل پدیده‌های غیرخطی و پیچیده، کاربردهای فراوانی در مسائل مختلف مهندسی آب از جمله رسوب پیدا کرده‌اند. با توجه به عدم قطعیت موجود در داده‌ها و فضای چند متغیری حاکم بر الگوی ورودی مدل‌ها، استفاده از روش‌های محاسبات نرم در محاسبات رسوب ورودی به آبخیز را اجتناب‌ناپذیر نموده است [۱۶]. لذا این مشکل ذهن محققین را به سمت روش‌هایی نوین فارغ از نیاز به معادلات مذکور سوق می‌دهد. در این میان می‌توان از روش هوش مصنوعی که مبتنی بر روابط حاکم بر طبیعت است و از قدرت انعطاف‌پذیری بالایی برخوردار است استفاده نمود. ناکاتو و همکاران [۱۲] جهت کنترل رسوب در آبخیز ایستگاه پمپاژ واحد ۳ نیروگاه کانسیل بلافس در رودخانه میسوری از صفحات مستغرق استفاده نمودند. آن‌ها به این

## استفاده از روش رگرسیونی و مدل GMDH در تخمین نسبت رسوب ورودی به آبخیزهای جانبی

امیر مرادی نژاد<sup>۱</sup>

تاریخ دریافت ۱۴۰۳/۰۳/۲۳ تاریخ پذیرش ۱۴۰۳/۰۶/۰۲

DOI: WMJI.2024.2032135.1075/10.22034

## چکیده

به علت تغییراتی که در توزیع سرعت در محدوده دهانه آبخیز رخ می‌دهد، معمولاً عمل رسوب‌گذاری صورت می‌گیرد که باعث کاهش راندمان آبخیزداری، افزایش هزینه‌های اجرایی برای عملیات رسوب‌زدایی و در نهایت تغییر مسیر و خط‌القدر رودخانه به سمت ساحل مقابل آبخیز می‌شود. استفاده از سازه‌های کنترل رسوب دیوار جداکننده در جلوی آبخیز و هم‌زمان آب‌شکن در مقابل آبخیز باعث کاهش رسوب ورودی و افزایش راندمان آبخیزداری می‌شود. در تحقیق حاضر تأثیر دیوار جداکننده و سازه آب‌شکن در تخمین نسبت رسوب ورودی به آبخیز به صورت آزمایشگاهی، روش‌های داده‌کاوی و رگرسیون چندگانه مورد ارزیابی قرار گرفته است. ابتدا با انجام آنالیز ابعادی، نسبت‌های بدون بعد استخراج و رابطه بین متغیرها و مقدار آن‌ها در آزمایش‌ها مشخص گردید. با استفاده از نرم‌افزارهای آماری XLSTAT و SPSS از روش گام‌به‌گام و رگرسیون استاندارد (اینتر) معادلاتی برای ارتباط بین متغیرهای مستقل و وابسته استخراج شد. بعد از به دست آوردن معادلات خطای نسبی هر معادله محاسبه شد. سپس بهترین معادله که  $R^2$  آن بالا و خطای نسبی آن پایین بود انتخاب و پیشنهاد شد. در مرحله بعد با روش‌های شبکه‌های عصبی مصنوعی روش کنترل گروهی داده‌ها (GMDH) مدل‌سازی انجام و بهترین روش در تخمین نسبت رسوب ورودی به آبخیز انتخاب شد. نتایج نشان داد که در تخمین نسبت رسوب ورودی به آبخیز بهترین عملکرد مربوط به مدل (GMDH) با شاخص‌های آماری  $MAPE=26/95$  و  $RMSE=0/039$ ،  $MAD=0/85$ ،  $R^2=0/03$  می‌باشد. در روش رگرسیون گام‌به‌گام  $RMSE=4/99$ ،  $R^2=0/38$  و در روش رگرسیون اینتر  $RMSE=4/16$ ،  $R^2=0/76$  می‌باشد.

۱- استادیار مرکز تحقیقات و آموزش کشاورزی و منابع طبیعی استان مرکزی،  
Email: amir\_24619@yahoo.com

نتیجه رسیدند که پس از ۳/۵ سال از زمان نصب صفحات مستغرق، مشکل رسوب گذاری آن برطرف شده است [۱۲]. ادگارد و وانگ [۱۴] نشان دادند آب‌شکن‌هایی که به سمت بالادست متمایل باشند، معمولاً موجب انحنای بیشتری در جریان اصلی می‌شوند [۱۴]. نیری و همکاران [۱۳] مدل عددی سه‌بعدی جریان روی انشعابی ۹۰ درجه در کانالی با مقطع مستطیلی را توسعه دادند و با نتایج آزمایشگاهی صحت‌یابی کرده‌اند. بر اساس یافته‌های این محققان، با افزایش نسبت انحراف جریان، عرض ناحیه گردابی کم و بر طول آن افزوده می‌شود. علاوه بر آن، برای نسبت دبی ۰/۶۴ علاوه بر گردابه داخل آبگیر، ناحیه چرخشی دیگری در داخل کانال اصلی و در مقابل دهانه آبگیر تشکیل می‌شود [۱۳]. بارکدول و همکاران [۲]، در تحقیقات خود نشان دادند که شدت رسوبات ورودی بستر به داخل آبگیر پس از نصب صفحات مستغرق زمانی می‌تواند ناچیز باشد که نسبت دبی واحد عرض آبگیر به دبی واحد عرض کانال اصلی (qr) کم‌تر از حدود ۰/۲ باشد. آن‌ها پس از بررسی آزمایشگاهی، برای افزایش qr و حفظ کارایی صفحات مستغرق، دو راه‌حل دیواره جانبی در کنار صفحات مستغرق و عریض کردن ورودی آبگیر را پیشنهاد دادند [۲]. ساجدی سابق و حبیبی [۱۷] تأثیر کاربرد توأم صفحات مستغرق و مجموعه آب‌شکن‌ها را در افزایش راندمان آبگیری بررسی کردند و نشان دادند با کاربرد توأم صفحات مستغرق و مجموعه آب‌شکن‌ها ورود رسوبات به آبگیر کاهش می‌یابد [۱۷]. رامامورتی و همکاران [۱۶] نشان دادند که با افزایش نسبت انحراف جریان، طول و عرض ناحیه جدایی جریان در کانال آبگیر کاهش می‌یابد. همچنین وسعت ناحیه جدایی در کانال آبگیر در کف کم‌تر از سطح است [۱۶]. گوهری و همکاران [۵] به کنترل رسوب در آبگیر با کاربرد صفحات مستغرق و آب‌شکن پرداختند، نتایج نشان داد که کارایی صفحات با احداث آب‌شکن در ضلع مقابل آبگیر افزایش پیدا کرده و مقدار ورود رسوبات به آبگیر کاهش یافته است. با جایجایی آب‌شکن به اندازه دو برابر عرض دهانه آبگیر به بالادست کانال اصلی، مقدار ورود رسوبات به آبگیر به‌طور متوسط ۴۰ درصد کاهش می‌یابد. همچنین ورود رسوبات به آبگیر با استفاده از آب‌شکن با زاویه ۹۰ درجه نسبت به دو آب‌شکن با زوایای ۴۵ و ۱۳۵ درجه به حداقل می‌رسد [۵]. سالم‌نیا و شفافی‌بجستانی [۱۸] اثر تغییر نسبت آبگیری بر میزان رسوب بستر ورودی به آبگیر در کانال ذوزنقه‌ای ( $z=1/5$ ) با زاویه آبگیری ۶۰ درجه نسبت به جهت جریان را با و بدون نصب صفحات مستغرق مورد بررسی قرار داد. نتایج آنان نشان داد که با افزایش نسبت آبگیری از ۷/۵ درصد به ۱۶ درصد، میزان رسوب ورودی به آبگیر در هر دو حالت با و بدون صفحات مستغرق به‌طور متوسط به میزان ۲۳ درصد افزایش می‌یابد [۱۸]. عطارزاده و همکاران [۱] به بررسی آزمایشگاهی تأثیر آستانه، آب‌شکن و صفحات مستغرق بر کنترل رسوب و توپوگرافی بستر در نسبت دبی آبگیری ۰/۱۲، ۰/۱۵ و ۰/۱۸ در آبگیرهای جانبی پرداختند. نتایج نشان داد در کنترل رسوب، تأثیر آب‌شکن بیش از سایر گزینه‌ها است [۱].

زمانی و همکاران [۲۰] در پژوهشی به بررسی آزمایشگاهی و عددی تأثیر مکان آب‌شکن بر شرایط هیدرولیکی و رسوبی آبگیر جانبی پرداختند. نتایج آن‌ها نشان داد که استفاده از آب‌شکن در پایین دست و روبروی آبگیر در ابعاد مناسب نه تنها می‌تواند باعث افزایش نسبت دبی انشعابی ورودی به آبگیر شود بلکه آب شستگی بستر کانال را هم کم می‌کند [۲۰].

مرادی‌نژاد و همکاران [۱۰] در مطالعه‌ای با استفاده از روش‌های مدل‌های شبکه عصبی مصنوعی، نروفازی، منحنی سنجه رسوب و همچنین مدل رگرسیونی چند متغیره به بررسی کارایی روش‌های برآورد بار رسوب معلق رودخانه قره‌چای پرداختند. نتایج آن‌ها نشان داد که روش نروفازی بر اساس ورودی‌های دبی و رسوب و مدل‌های شبکه عصبی مصنوعی بر اساس ورودی‌های دبی، نسبت به رگرسیون چند متغیره و منحنی سنجه رسوب، دارای دقت مطلوب‌تری می‌باشند [۱۰]. بیرانوند و همکاران [۳] در پژوهشی با استفاده از الگوریتم‌های یادگیری ماشین به بررسی کارایی مدل‌های GP-RBF، GP-PUK، RF، RepTree، M5P در مدل‌سازی بار معلق رودخانه در دوره‌های کم‌آبی و پرآبی در حوزه آبخیز کشکان پرداختند. نتایج آن‌ها نشان داد با توجه به معیارهای ارزیابی مدل GP با دو کرنل PUK و RBF در دوره کم‌آبی و پرآبی عملکرد بهتری را نسبت به سایر مدل‌ها (RF، RepTree، M5P) داشته است. همچنین با توجه به نتایج بخش آزمایش مدل GP-PUK بهترین نتیجه را داشته است [۳]. کشتکار و همکاران [۷] در تحقیقی از روش‌های نرم برای تخمین رسوب معلق رودخانه در پاکستان پرداختند. آن‌ها نتایج مدل‌های RM5Tree با مدل‌های رگرسیون بردار پشتیبان (SVR)، شبکه عصبی مصنوعی (ANN)، اسپلاین رگرسیون تطبیقی چند متغیره (MARS)، منحنی سنجه رسوب (SRC) و روش سطح پاسخ (RSM) مقایسه کردند. از پارامترهای دبی جریان، پوشش برف، دمای هوا، تبخیر و تعرق و بارندگی مؤثر استفاده کردند. دقت حاصل از مدل‌ها با استفاده از ضریب همبستگی پیرسون ( $R^2$ )، ریشه میانگین مربع خطا (RMSE) و میانگین درصد مطلق خطا (MAPE) ارزیابی شد. نتایج نشان داد که مدل RM5Tree پیک‌های رسوب معلق را با دقت نسبی ۸۴/۱۰ درصد در مقایسه با مدل‌های ANN، MARS، SVR، M5Tree، RSM و SRC به ترتیب با دقت ۸۰/۶۲، ۷۷/۸۶، ۸۱/۹۰، ۸۰/۵۸، ۸۰/۵۸ و ۷۶ درصد بهتر پیش‌بینی کرده است [۵]. مرادی‌نژاد و همکاران [۱۱] در مطالعه‌ای به عملکرد سه نوع مدل ماشین بردار پشتیبان (SVM)، سیستم عصبی فازی-تطبیقی (ANFIS) و برنامه‌ریزی بیان ژن (GEP) در شبیه‌سازی بار رسوبی رودخانه‌ها پرداختند. بررسی و نتایج به‌دست‌آمده از این سه روش را با یکدیگر و با نتایج منحنی سنجه مقایسه کردند. نتایج بیانگر عملکرد قابل قبول مدل‌ها نسبت به منحنی سنجه رسوب بود. در بین سه مدل، مدل (GEP) با بیش‌ترین ضریب تعیین  $R^2$  با مقدار ۰/۹۹ و کم‌ترین ریشه میانگین مربعات خطا RMSE برحسب تن در روز با مقدار ۰/۱۰ نسبت به دو مدل دیگر برتری داشت [۱۱]. محسنی [۹] در مطالعه‌ای

(۲)

$$F\left(\frac{V_m}{\sqrt{gd}}, \frac{Q_L}{\sqrt{gd^5}}, \frac{Q_m}{\sqrt{gd^5}}, \frac{Q_{SL}}{\sqrt{gd^5}}, \frac{Q_{sm}}{\sqrt{gd^5}}, \frac{k_s}{d}, \frac{B_m}{d}, \frac{d_{50}}{d}, \frac{\rho_s}{\rho}, \frac{v}{d\sqrt{gd}}, t\sqrt{\frac{g}{d}}, \frac{b}{d}, \frac{L}{d}, \frac{L_1}{d}, \frac{L_2}{d}, \frac{L_D}{d}, \frac{H}{d}, \beta_1, \beta_2, \alpha, \delta, S_L, S_m\right) = 0$$

با انجام عملیات جبری روی نسبت‌های بدون بعد معادله (۲)، نسبت‌های بدون بعد تغییر کرد. سپس با توجه به ثابت بودن  $B_m, \delta$ ، حذف نسبت‌های  $\beta_2, b, \alpha, L_2, L_1, L_D, K_s, S_L, S_m, t, v, \rho, \rho_s, d_{50}, g$ ، رابطه بدون بعد نهایی به صورت معادله (۳) قابل ارائه می‌باشد.

$$G_r = f(Fr, Q_r, \frac{H}{d}, \beta_1) \quad (3)$$

نسبت  $G_r = \frac{Q_d}{Q_m}$  نسبت دبی رسوب ورودی به کانال آبیگر به دبی رسوب بالادست کانال اصلی،  $Q_r = \frac{Q_L}{Q_m}$  نسبت دبی کانال آبیگر به دبی کانال اصلی (نسبت آبیگری)،  $Fr$  عدد فرود جریان بالادست آبیگر،  $\beta$  زاویه بین دو ضلع سازه دیوار جداکننده،  $\frac{H}{d}$  نسبت ارتفاع خارج از بستر دیوار جداکننده به عمق جریان در کانال اصلی می‌باشد.

### تجهیزات آزمایشگاهی و روش انجام آزمایش‌ها

آزمایش‌ها بر روی فلومی به طول ۱۲ متر و عرض ۱/۵ و ارتفاع ۰/۹ متر و دارای سیستم گردش آب و رسوب در پژوهشکده تحقیقات حفاظت خاک و آبخیزداری انجام گرفت. آبیگری توسط کانال جانبی به عرض ۰/۶ متر و طول ۲/۵ و با زاویه ۶۰ درجه نسبت به جهت جریان در کانال اصلی انجام شده است. کانال اصلی دارای شیب ۰/۰۲، آبیگر افقی و تراز آن هم‌سطح با بستر رسوبی کانال اصلی بوده است. طول آب‌شکن بکار رفته  $B_m/25$ ، فواصل آب‌شکن از مقابل مرکز آبیگر (۲b, ۲b, b/۳b, ۵) و زوایای آن با جهت جریان (۳۰، ۶۰، ۹۰) درجه بوده است.

برای اندازه‌گیری جریان در کانال‌های اصلی و آبیگر از سرریزهای لبه تیز مستطیلی و مثلثی، برای اندازه‌گیری نیمرخ سطح آب از ارتفاع سنج<sup>۱</sup> و بستر نگار<sup>۲</sup> با دقت ۰/۱ ± میلی‌متر استفاده گردید. شکل (۲) نمایی از فلوم به کار برده شده در آزمایش‌ها، موقعیت دیوار جداکننده و آب‌شکن را نشان می‌دهد. جدول (۱) مقادیر داده‌های (بیش از ۷۰ درصد) موردنیاز برای آموزش و بخشی هم (معمولاً بین ۲۰ تا ۳۰ درصد) برای اعتبار سنجی استفاده می‌شوند. محدوده تغییرات و مشخصات آماری پارامترهای در جدول (۱) ارائه شده است.

به پیش‌بینی سرعت متوسط عمقی در کانال مرکب با پوشش گیاهی در سیلاب‌دشت با استفاده از مدل برنامه‌ریزی بیان ژن پرداخت. نتایج ارزیابی عملکرد مدل‌ها بر مبنای شاخص‌های عملکرد نشان داد که مدل ارائه‌شده توسط GEP با ضریب همبستگی ۰/۹۰ و ۰/۸۴ و ریشه میانگین مربعات خطا ۰/۰۸۳ و ۰/۰۸۳ به ترتیب، در کانال اصلی و سیلاب‌دشت، در پیش‌بینی سرعت به‌خوبی عمل نموده و در موارد عملی قابل اجرا است [۹]. با توجه به تحقیقات انجام‌شده ملاحظه می‌شود که در مورد نسبت رسوب ورودی به آبیگر در حالت وجود سازه‌های کنترل رسوب ترکیب دیوار جداکننده و آب‌شکن و روش‌های داده‌کاوی تحقیقاتی انجام‌نشده است. لذا هدف از این پژوهش ارزیابی و مقایسه روش‌های داده‌کاوی و رگرسیون چند متغیره در برآورد نسبت بار رسوب ورودی به آبیگر و مقایسه با نتایج آزمایشگاهی می‌باشد. از آنجائی که ارزیابی روش‌های مبتنی بر داده‌کاوی به‌عنوان جایگزین برای تخمین بار رسوب ورودی به آبیگر جانبی با ترکیب سازه دیوار جداکننده تاکنون استفاده نشده است. بدین منظور به عملکرد سه نوع مدل در شبیه‌سازی بار رسوبی به آبیگر پرداخته، سپس نتایج سه روش با یکدیگر و با نتایج رگرسیونی و آزمایشگاهی مورد مقایسه قرار گرفت.

### مواد و روش‌ها

#### تحلیل ابعادی

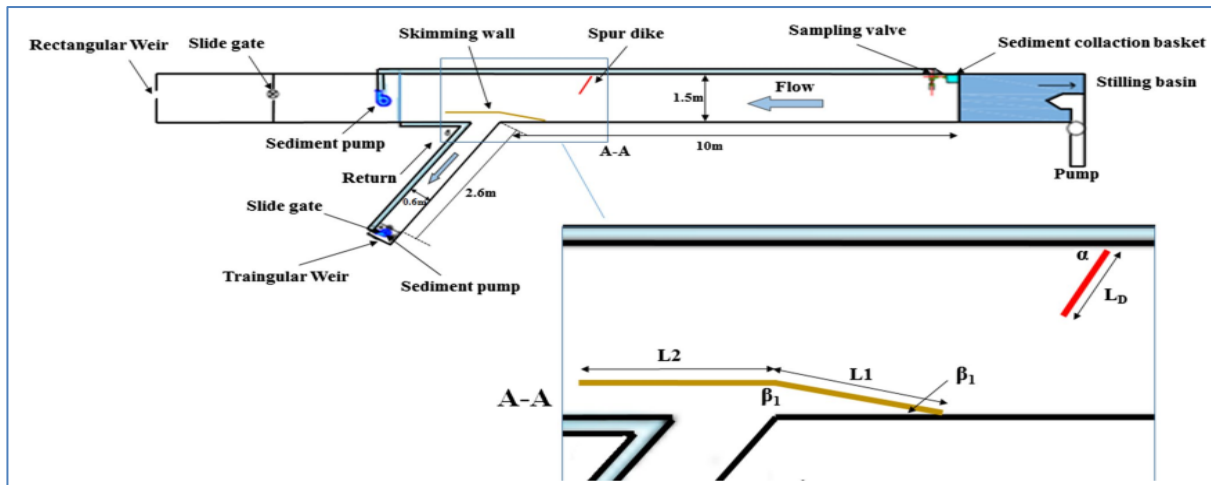
پارامترهای زیادی در جریان ورودی به آبیگر مؤثر می‌باشند، این پارامترها عبارت‌اند از دبی جریان در کانال اصلی ( $Q_m$ )، دبی جریان در کانال آبیگر ( $Q_d$ )، دبی رسوب در کانال اصلی ( $Q_{sm}$ )، دبی رسوب در کانال آبیگر ( $Q_{sd}$ )، عمق جریان در کانال اصلی ( $d$ )، سرعت جریان در کانال اصلی ( $V_m$ )، عرض کانال اصلی ( $B_m$ )، عرض کانال آبیگر ( $b$ )، شیب کانال اصلی ( $S_m$ )، شیب کانال فرعی ( $S_L$ )، شتاب ثقل ( $g$ )، دانسیته جریان ( $\rho$ )، لزجت سینماتیکی سیال ( $\nu$ )، زاویه کانال آبیگر با کانال اصلی ( $\delta$ )، زاویه آب‌شکن با ساحل کانال اصلی ( $\alpha$ )، طول آب‌شکن ( $L_D$ )، فاصله آب‌شکن از خط مرکزی کانال آبیگر ( $L$ )، زاویه سازه دیوار جداکننده با ساحل ( $\beta_1$ )، زاویه بین دو ضلع سازه دیوار جداکننده ( $\beta_2$ )، ارتفاع صفحات دیواره جداکننده ( $H$ )، طول شاخه ابتدایی دیوار جداکننده ( $L_1$ )، طول شاخه دوم دیوار جداکننده ( $L_2$ )، زمان آزمایش ( $t$ )، قطر متوسط رسوبات ( $d_{50}$ )، دانسیته رسوبات ( $\rho_s$ )، ارتفاع زبری رسوبات بستر ( $K_s$ )، تعداد ۲۶ پارامتر دخیل در مسئله به صورت تابع (۱) در آمد.

(۱)

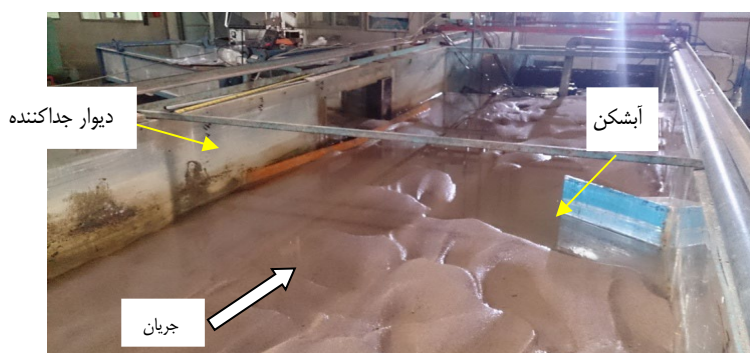
$$n, Q_{SL}, Q_{sm}, b, L, H, d, B_m, V_m, L_1, L_2, L_D, S_L, S_m, K_s, d_{50}, \rho_s, \rho, \beta_1, \beta_2, \alpha, \delta, t, \nu, g = 0$$

با استفاده از تحلیل ابعادی و روش باکینگهام نسبت‌های بدون بعد تعیین شد و معادله (۱) به صورت (۲) درآمد.

1.Point Gage  
2.Bed Profiler



شکل ۱: نمای کلی فلوم، آب شکن، دیوار جداکننده و سیستم چرخش آب و رسوب



شکل ۲: نمای فلوم، آب شکن، دیوار جداکننده جلوی دهانه آبگیر

که  $X$  نشان دهنده متغیر ورودی،  $y$  متغیر خروجی،  $n$  تعداد ورودی‌ها و  $a$  ضریب است. سری Volterra که یک شکل کلی معادله ریاضی است و می‌تواند با استفاده از چند جمله‌ای درجه دوم مانند معادله (۵) تقریب زده شود. در الگوریتم GMDH در محیط نرم افزار متلب برای تولید نرون‌های لایه اول تمام متغیرهای ورودی باهم ترکیب شده و تمام ترکیبات دوتایی متغیرهای ورودی در نظر گرفته می‌شوند. به عنوان مثال اگر تعداد متغیرهای ورودی برابر  $n$  باشد، تعداد نرون‌های لایه اول برابر با  $L1 = \binom{n}{2}$  محاسبه می‌شود که در این رابطه  $n$  تعداد متغیرهای ورودی است [۶].

$$G(X_1, a_1 X_1^2 + a_2 X_2^2 + a_3 X_1 + a_4 X_2 + a_5 X_1 X_2 + a_6 \quad (5)$$

$$\bar{y} = X_j =$$

پارامترهای مجهول در معادله (۵) ضرایب چند جمله‌ای است که با استفاده از روش‌های رگرسیون به گونه‌ای محاسبه می‌شود که اختلاف بین خروجی واقعی  $y$  و مقادیر محاسبه شده برای هر جفت متغیرهای ورودی به حداقل برسد. ضرایب شش گانه هر نرون در شبکه از طریق رویکرد حداقل مربعات محاسبه می‌شوند. در طراحی روش GMDH، هدف جلوگیری از رشد واگرایی شبکه و نیز مرتب کردن شکل و ساختار شبکه به یک یا چند پارامتر عددی بوده، به گونه‌ای که با تغییر این پارامتر ساختار شبکه نیز تغییر می‌کند.

جدول ۱: آمار خلاصه داده‌های کمی (داده‌های کمی)

متغیرها	تعداد داده‌ها	کمینه	بیشینه	متوسط	انحراف معیار
$Q_r$	۱۱۲	۰/۰۹۶	۰/۲۸۹	۰/۱۵۰	۰/۰۳۵
$\beta$	۱۱۲	۷	۱۸	۱۳/۴۶	۴/۲۸
$H/d$	۱۱۲	۰/۰۱۶	۰/۰۹	۰/۰۴۹	۰/۰۲۳
$F_r$	۱۱۲	۰/۰۰۲	۰/۲۴۶	۰/۰۳۱	۰/۰۴۱
$G_r$	۱۱۲	۰/۰۳۶	۰/۵۲۴	۰/۱۴۱	۰/۱۰۲

### روش گروهی مدل‌سازی داده‌ها (GMDH)

ایواخنکو (۱۹۷۶) اولین بار الگوریتم GMDH را به عنوان روشی برای شناسایی روابط غیرخطی بین متغیرهای ورودی و خروجی پیشنهاد داد [۶]. این الگوریتم یک رویکرد خودسازمانی بوده که توانایی به کارگیری در مدل‌سازی سامانه‌های پیچیده را دارد. شبکه عصبی GMDH دارای ساختار شبکه‌ای چندلایه و پیش‌خور و مجموعه‌ای از نرون‌ها است. اتصال بین متغیرهای ورودی و خروجی را می‌توان با استفاده از سری چند جمله‌ای پیچیده به نام سری Volterra به صورت معادله (۴) بیان کرد.

$$\bar{y} = a_0 + \sum_{i=1}^n a_i X_i + \sum_{i=1}^n \sum_{j=1}^n a_{ij} X_i X_j + \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n a_{ijk} X_i X_j X_k \dots$$

1. Group Method of Data Handling

## رگرسیون هم‌زمان و گام‌به‌گام

رگرسیون یک روش آماری برای سنجش تأثیرات متغیرها با در نظر گرفتن روابط هم‌زمان آن‌ها بر یکدیگر است. در واقع، بررسی اثر چند متغیر مستقل بر متغیر وابسته است. در این تحقیق از روش رگرسیون هم‌زمان و گام‌به‌گام استفاده شده است. در روش رگرسیون هم‌زمان، تمام متغیرهای مستقل باهم وارد تحلیل می‌شوند. ولی در روش گام‌به‌گام، متغیرها را یک‌به‌یک وارد مدل می‌کند؛ یعنی ابتدا متغیری که بالاترین ضریب همبستگی را با متغیر وابسته دارد، وارد تحلیل می‌شود. در این روش ترتیب ورود متغیرها در دست محقق نیست تمامی متغیرهای مستقل وارد مدل می‌شوند و آن متغیر مستقلی که تأثیر چندانی بر متغیر وابسته نداشته باشد از مدل حذف می‌شود. میزان اهمیت متغیر در مدل، مقدار سطح معنی‌داری یا آماره  $F$  متناظر با آن در جدول آزمون معنی‌داری متغیرها است. اگر متغیری  $F$  بیش‌تری از  $F$  در نظر گرفته شده داشته باشد وارد مدل نمی‌شود. در این تحقیق برای تشخیص اهمیت متغیر از روش گام‌به‌گام استفاده شده است.

## نتایج و بحث

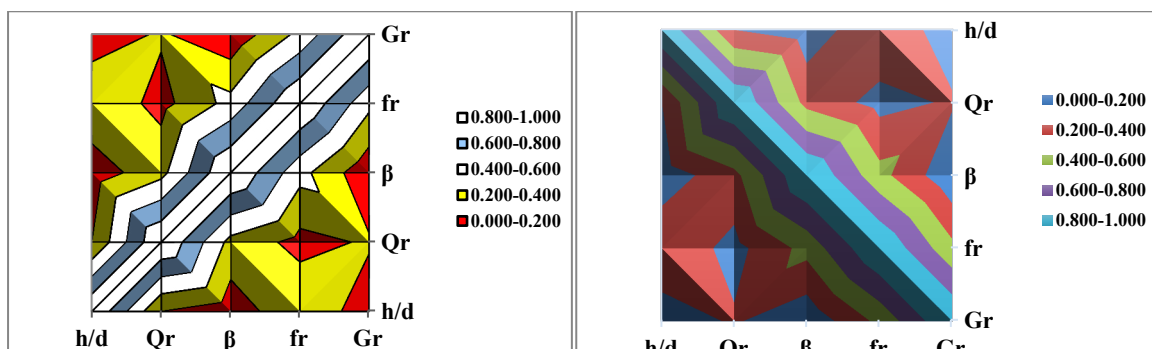
به‌منظور مدل‌سازی تخمین نسبت رسوب ورودی به آبگیر در ابتدا بایستی همبستگی بین داده‌های ورودی مدل مورد بررسی قرار گیرد. آزمون نرمالیته برای داده‌ها با انجام آزمون‌هایی (آزمون‌های شاپیرو ویلک، اندرسون دارلینگ، لیلیه فورس و جارکویرا) توسط نرم‌افزار آماری XLSTAT مورد بررسی قرار گرفته و نتایج مربوطه در جدول (۲) ارائه گردیده است. برای بررسی اینکه داده‌ها دارای توزیع نرمال هستند یا نه از آزمون‌های شاپیرو ویلک، اندرسون دارلینگ، لیلیه فورس و جارکویرا استفاده شده است. در این آزمون‌ها فرض صفر برابر با نرمال بودن داده‌ها و فرض مقابل برابر با غیرنرمال بودن داده‌ها در نظر گرفته شده است. همان‌طور که مشاهده می‌گردد داده‌ها در تمامی آزمون‌های مورد بررسی نرمال بوده است. به کمک این آزمون‌ها می‌توان راجع به میزان انحراف داده‌ها از نرمال و اثر آن بر عملکرد رگرسیون‌گیری اظهار نظر کرد. با توجه به اهمیت استفاده از داده‌های صحیح آماری، کلیه داده‌های موجود از نظر همگن بودن

توسط آزمون همگنی نرمال استاندارد که یکی از روش‌های متداول برای ارزیابی همگنی داده‌ها می‌باشد، مورد بررسی قرار گرفت. شکل (۴) ماتریس همبستگی داده‌ها را نشان می‌دهد. جدول (۲) مقدار و ضرایب همبستگی داده‌های آزمایشگاهی را نشان می‌دهد. در جدول (۳) همبستگی بین داده‌های ورودی نسبت به پارامترهای آورده شده است. سری زمانی و فراوانی داده‌ها، پارامترهای مستقل و وابسته بی‌بعد در شکل‌های (۵) تا (۸) ارائه شده است. با توجه به شکل‌ها، مقادیر پارامترهای بی‌بعد مستقل و پارامتر بی‌بعد رسوب ورودی به آبگیر از توزیع نرمال تبعیت می‌کنند. لازم است گفته شود در روش‌های داده‌کاوی لزومی به تبعیت پارامترهای به‌کاررفته در مدل‌سازی از توزیع نرمال نیست و از این‌روی مشکلی در فرآیند مدل‌سازی ایجاد نمی‌کند. به‌عنوان یک قانون کلی، در فن‌های آماری فرض اولیه بر این اساس است که توزیع داده‌ها مشخص است که بیش‌تر موارد فرض بر این است که توزیع نرمال است و در نهایت درستی یا نادرستی نتایج نهایی به درست بودن فرض اولیه وابسته است. در مقابل، روش‌های داده‌کاوی از هیچ فرض در مورد داده‌ها استفاده نمی‌کنند [۱۹ و ۸].

جدول ۲: مقدار و ضریب همبستگی داده‌های آزمایشگاهی

متغیرها	h/d	Qr	$\beta$	fr	Gr
h/d	۱	۰/۰۳۹	۰/۰۰۶	۰/۱۳۳	۰/۰۰۸
Qr	۰/۰۳۹	۱	۰/۱۱۶	۰/۰۱۸	۰/۰۵۰
$\beta$	۰/۰۰۶	۰/۱۱۶	۱	۰/۲۲۱	۰/۰۰۲
fr	۰/۱۳۳	۰/۰۱۸	۰/۲۲۱	۱	۰/۲۰۵
Gr	۰/۰۰۸	۰/۰۵۰	۰/۰۰۲	۰/۲۰۵	۱

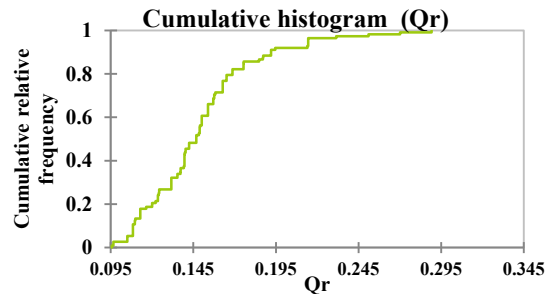
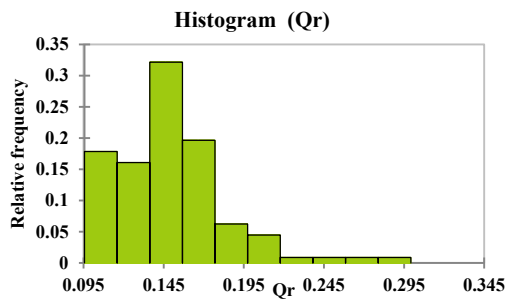
در این تحقیق عملکرد الگوریتم هوشمند GMDH و رگرسیون چندگانه جهت پیش‌بینی مقادیر رسوب ورودی به آبگیر جانبی با استفاده از داده‌های آزمایشگاهی مورد مقایسه قرار گرفت. جهت تخمین رسوب ورودی به آبگیر ابتدا داده‌های پارامترهای بی‌بعد، نسبت جریان ورودی به آبگیر، عدد فرود، زاویه دیوار جداکننده و نسبت ارتفاع خارج از بستر دیوار جداکننده به عمق جریان



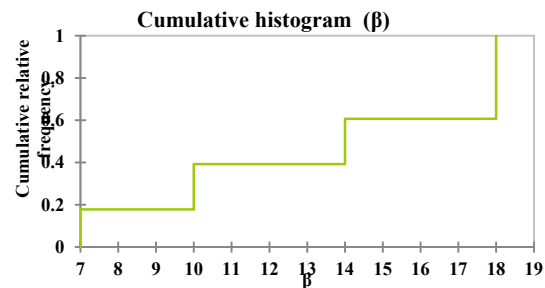
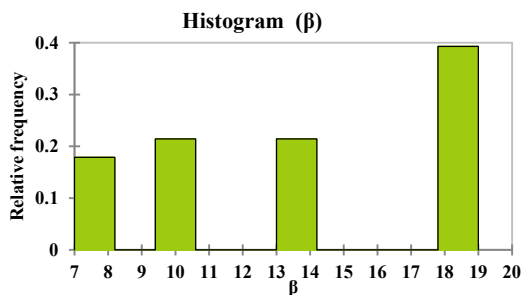
شکل ۴: ماتریس همبستگی داده‌های آزمایشگاهی

جدول ۳: نتایج آزمون‌های تعیین نرمال بودن متغیرها

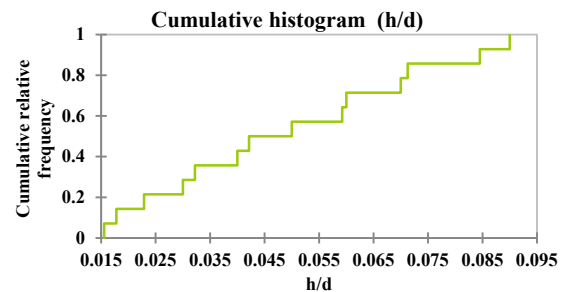
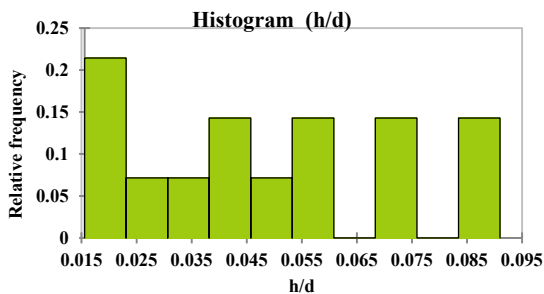
Test	Shapiro-Wilk		Anderson-Darling		Lilliefors		Jarque-Bera	
	p-value (Two-tailed)	alpha	p-value (Two-tailed)	alpha	p-value (Two-tailed)	alpha	p-value (Two-tailed)	alpha
$Q_r$	</0.0001	0.05	</0.0001	0.05	</0.0001	0.05	</0.0001	0.05
$\beta$	</0.0001	0.05	</0.0001	0.05	</0.0001	0.05	</0.0001	0.05
H/d	</0.0001	0.05	</0.0001	0.05	0	0.05	0.031	0.05
$F_r$	</0.0001	0.05	</0.0001	0.05	</0.0001	0.05	</0.0001	0.05
$G_r$	</0.0001	0.05	</0.0001	0.05	</0.0001	0.05	</0.0001	0.05



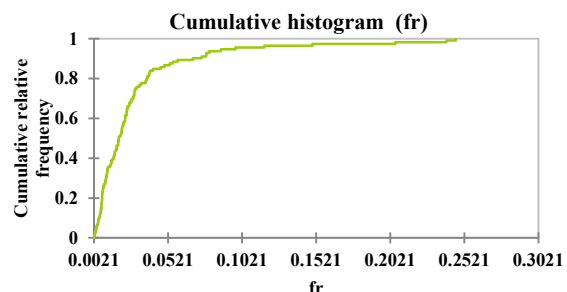
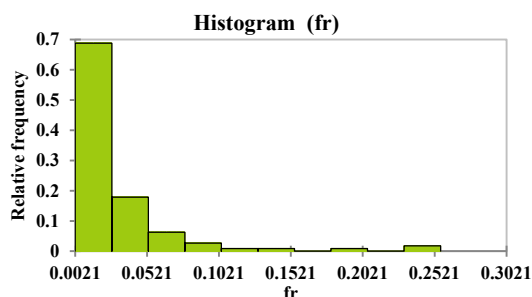
شکل ۵: هیستوگرام و فراوانی داده‌های بدون بعد ( $Q_r$ )



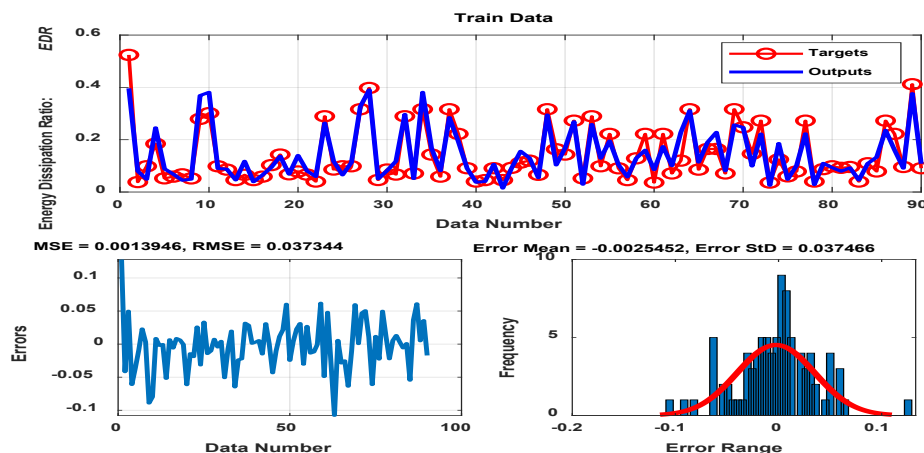
شکل ۶: هیستوگرام و فراوانی داده‌های بدون بعد ( $\beta$ )



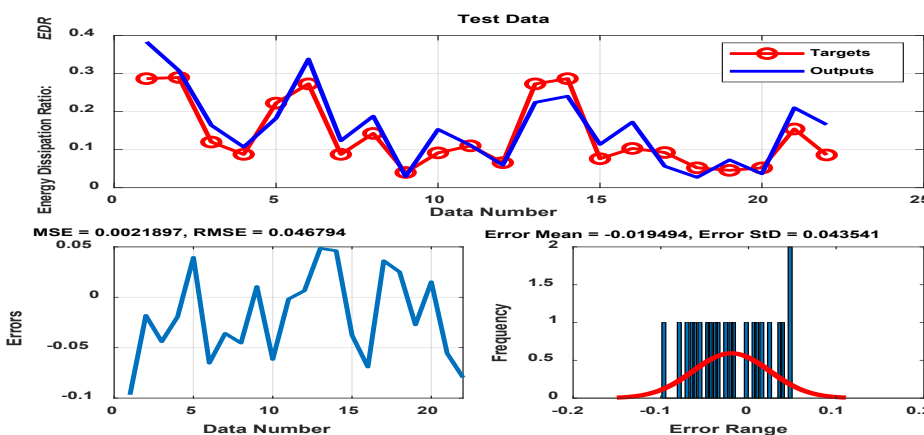
شکل ۷: هیستوگرام و فراوانی داده‌های بدون بعد ( $h/d$ )



شکل ۸: هیستوگرام و فراوانی داده‌های بدون بعد ( $Fr$ )



شکل ۹: نتایج خروجی مدل (GMDH) در مرحله آموزش



شکل ۱۰: نتایج خروجی مدل (GMDH) در مرحله آزمون

استفاده از همبستگی خطی چندمتغیره با دو روش گام به گام<sup>۱</sup> و اینتر<sup>۲</sup> مورد بررسی قرار گرفت. در روش گام به گام متغیرهای کم اهمیت چون F آن‌ها بیش از F در نظر گرفته شده بود وارد مدل نشدند. متغیرهای دیگر که بیشترین همبستگی را با متغیر وابسته دارند وارد مدل شد. در این رابطه سعی شده نسبت انحراف رسوب به آبگیر به صورت تابعی از پارامترهای بی بعد باشد. برای بررسی هر کدام از روابط به دست آمده از تحلیل همبستگی، خطای نسبی هر رابطه محاسبه گردید و رابطه بهتر بر اساس معیار خطای نسبی و  $R^2$  انتخاب شده است. برای تجزیه و تحلیل داده‌های به دست آمده از آزمایش‌ها، از نرم افزار آماری (SPSS) و (xlstat) استفاده گردید، داده‌ها وارد نرم افزار شدند، سپس پارامترهای وابسته و مستقل مشخص شدند. پارامترهای مستقل ۴ پارامتر و پارامتر وابسته فقط  $(G_p)$  بود. برای استفاده از رگرسیون چندمتغیره خطی باید ابتدا مدل ریاضی به نرم افزار داد تا بتواند ضرایب آن را محاسبه کند. نتایج رگرسیون چندگانه منجر شد به دو معادله که بهترین نتیجه را دارند. معادله (۶) از روش گام به گام و معادله (۷) از روش اینتر به دست آمده است. نتایج شاخص‌های هر دو معادله در جدول (۴)

به عنوان پارامتر مستقل و نسبت رسوب ورودی به آبگیر به عنوان پارامتر وابسته انتخاب و استفاده شد. سپس تأثیر هر یک از این متغیرها در مدل سازی نسبت رسوب ورودی به آبگیر مشخص شد. جهت به کارگیری این مدل، از برنامه تدوین شده در محیط نرم افزار MATLAB استفاده شد. پس از مرتب کردن داده‌ها و تعیین پارامترهای مستقل و وابسته، داده‌ها وارد مدل شدند و مدل اجرا گردید. در این تحقیق ۷۰ درصد داده‌ها برای آموزش و ۳۰ درصد برای آزمون مورد استفاده قرار گرفت. نتایج خروجی مدل (GMDH) در مرحله آموزش و آزمون در اشکال (۹) و (۱۰) آمده است. با توجه به شکل در مرحله آموزش مقادیر شاخص‌های  $R^2 = 0.86$ ،  $MAD = 0.03$ ،  $MSE = 0.0014$ ،  $RMSE = 0.037$ ،  $RMSE = 0.022$ ،  $MAD = 0.032$ ،  $R^2 = 0.83$ ،  $MSE = 0.046$  به دست آمد. با توجه به شکل‌ها نتایج مدل سازی با این روش رضایت بخش می باشد.

رگرسیون چندگانه: پس از بررسی رابطه بین نسبت انحراف رسوب به آبگیر، به عنوان متغیر وابسته و تک تک پارامترها به عنوان متغیر مستقل، رابطه بین پارامترهای بی بعد به عنوان متغیر مستقل و نسبت انحراف رسوب به آبگیر  $(G_p)$  به عنوان متغیر وابسته با

1. Stepwise  
2. Enter

کننده مقدار  $G_r$  استفاده نمود. مدل رگرسیونی توانسته است، نسبت به حالت تصادفی، واریانس متغیر وابسته را بیش تر توصیف کند. از معادله های به دست آمده می توان به عنوان یک مدل پیش بینی کننده مقدار  $G_r$  استفاده نمود. نمودارهای دوطرفه فراوانی و اثرات باقیمانده خطای رگرسیون نشانگر و نمودار دوطرفه p-p plot حاکی از نرمال بودن داده ها دارد (شکل ۱۱ و ۱۲).

$$G_r = -0.103 + 0.322 \left(\frac{H}{d}\right) + 0.741F_r + 0.368\beta \quad (6)$$

$$G_r = -0.143 + 0.3 \left(\frac{H}{d}\right) + 0.73F_r + 0.31\beta + 0.152Q_r \quad (7)$$

جدول ۴: نتایج شاخص های معادلات (۶) و (۷)

شاخص	$R^2$	SE
روش گام به گام	۰/۳۸	۰/۰۸۱
روش اینتر	۰/۷۶	۰/۰۸

آمده است. در روش گام به گام متغیرهای  $Fr$ ,  $\beta$ ,  $H/d$  وارد معادله رگرسیون شدند ضریب احتمال برای ورود متغیرهای مستقل،  $0/50$  و ضریب احتمال برای خارج کردن متغیرها از معادله،  $0/1$  در نظر گرفته شده است. در مدل رگرسیونی به روش گام به گام،  $Fr$  نخستین متغیری بود که وارد مدل شد و به تنهایی  $19/8$  درصد از تغییرات  $G_r$  را توجیه نمود. پس از آن  $\beta$  وارد مدل شد و به همراه  $G_r$   $28$  درصد از تغییرات را توجیه نمود. در مرحله بعد  $H/d$  نیز به مدل اضافه شد و با دو متغیر دیگر در مجموع  $36$  درصد تغییرات  $G_r$  را توجیه نمودند. در بخش آنالیز واریانس (ANOVA) می توان گفت که مدل رگرسیونی توانسته است، نسبت به حالت تصادفی، واریانس متغیر وابسته را بیش تر توصیف کند، زیرا  $sig = 0$  کم تر از احتمال خطای اول  $\alpha = 0/05$  است. طبق جدول (۵) مقدار ثابت یا همان «عرض از مبدأ» برابر با  $-0/103$  و «شیب خط» نیز در قالب مقادیر بتا (Beta) به دست آمده است. با توجه به اینکه درجه آزادی در مقادیر بتا آمده است معادله خط بر اساس مقادیر بتا نوشته می شود؛ بنابراین معادله خط به صورت معادله (۶) می باشد؛ که در این معادله کلیه پارامترها قبلاً تعریف شده اند. از این معادله می توان به عنوان یک مدل پیش بینی

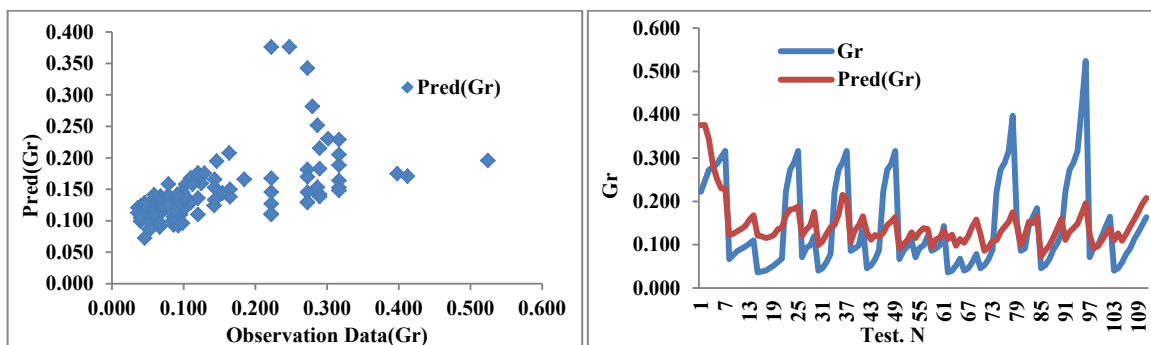
جدول ۵: نتایج تجزیه واریانس

Coefficients <sup>a</sup>						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	۰/۱۰۷	۰/۰۱۱	-	۹/۹۴۷	۰
	Fr	۱/۱۱۵	۰/۲۰۹	۰/۴۵۳	۵/۳۳۲	۰
2	(Constant)	-۰/۰۱۱	۰/۰۳۴		-۰/۳۰۹	۰/۷۵۸
	Fr	۱/۴۹۶	۰/۲۲۵	۰/۶۰۸	۶/۶۵۳	۰
	$\beta$	۰/۰۰۸	۰/۰۰۲	۰/۳۳۰	۳/۶۰۷	۰
3	(Constant)	-۰/۱۰۳	۰/۰۴۰	-	-۲/۵۹۴	۰/۰۱۱
	Fr	۱/۸۲۳	۰/۲۲۷	۰/۷۴۱	۸/۰۳۶	۰
	$\beta$	۰/۰۰۹	۰/۰۰۲	۰/۳۶۸	۴/۲۵۶	۰
	H/d	۱/۴۴۶	۰/۳۶۶	۰/۳۲۲	۳/۹۴۷	۰

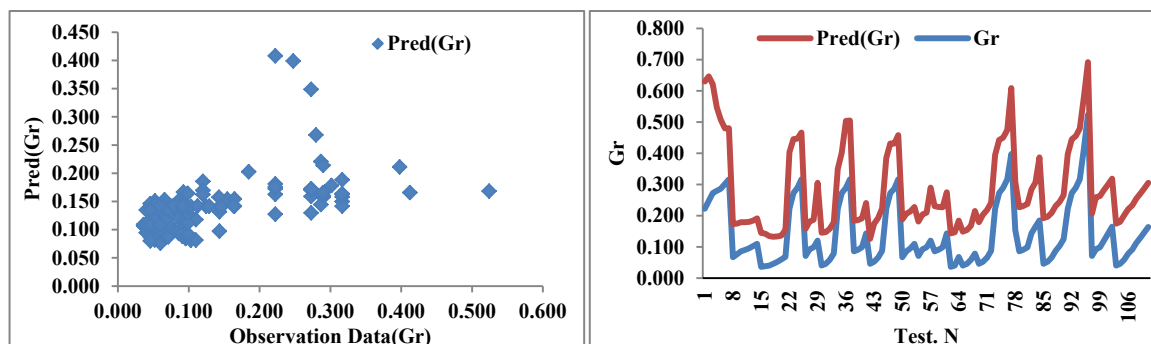
a. Dependent Variable: Gr

جدول ۶: مقایسه عملکرد مدل های مورد استفاده جهت تخمین نسبت رسوب ورودی به آبیگر

مدل	$R^2$	RMSE	MSE	MAD	MAPE
GMDH	۰/۸۵	۰/۰۳۹	۰/۰۰۲	۰/۰۳	۲۶/۹۵
رگرسیون گام به گام	۰/۳۸	۴/۹۹	۲۴/۹۶	۴/۷۴	۵۲۰۵
رگرسیون اینتر	۰/۷۶	۴/۱۶	۱۷/۳۲	۳/۹۵	۴۳۳۵



شکل ۱۱: خروجی از روش مدل رگرسیون گام به گام



شکل ۱۲: خروجی رگرسیون چندگانه از روش اینتر

جانبی و حفاظت از سواحل رودخانه به کار می‌رود. در جلوی دهانه آبرگیر جانبی در کانال اصلی و یا داخل رودخانه نصب می‌شود. این سازه در داخل جریان مستغرق می‌باشد و مقدار ارتفاع خارج از بستر دیوار سازه به میزان تقریباً (۲۵-۰/۵) درصد عمق آب در کانال اصلی و یا یک سوم عمق جریان در نظر گرفته می‌شود. البته بر اساس عمق جریان، دانه‌بندی رسوبات و توپوگرافی بستر این ارتفاع می‌تواند تغییر کند. سازه Skimming wall، می‌تواند از دوشاخه در جلوی آبرگیر تشکیل شود. یکی از شاخه‌ها ( $L_1$ ) از یک طرف به شاخه دوم ( $L_2$ ) به صورت مورب و از طرف دیگر به دیوار ساحل وصل می‌باشد. شاخه دیگر ( $L_2$ ) به صورت موازی با جریان می‌باشد که یک طرف آن به شاخه اول وصل می‌باشد، طرف دیگر آن موازی با جریان در کانال اصلی تا انتهای دهانه آبرگیر ادامه دارد شکل (۳). بر اساس نظر بارکدول و همکاران [۲] نسبت طول شاخه‌ها  $\frac{L_2}{L_1} = 1/5$  می‌باشد. زاویه نصب سازه به لبه کناری رودخانه ( $\beta_1$ ) باید کم‌تر از ۲۰ درجه باشد تا به‌عنوان مانع در جلوی جریان مشکل ایجاد نکند. زاویه بین دوشاخه سازه ( $\beta_2$ )، زاویه بین سازه با ساحل ( $\beta_1$ )، ابعاد سازه شامل طول و ارتفاع خارج از بستر صفحات بستگی به مشخصات رودخانه و مشخصات هندسی و هیدرولیکی جریان مانند عمق جریان و سرعت جریان دارد. برای به دست آوردن ابعاد دقیق سازه باید از مدل‌های فیزیکی و هیدرولیکی و یا عددی استفاده شود تا بتوان متناسب با سازه آبرگیر موردنظر، ابعاد و زوایای دیوار جداکننده را با بهترین راندمان به دست آورد. معیار بهترین راندمان، یعنی بیش‌ترین دبی با کم‌ترین رسوب وارد آبرگیر شود.

پس از مدل‌سازی داده‌ها توسط مدل GMDH و رگرسیون چندگانه خطی، مقدار نسبت رسوب ورودی به آبرگیر توسط این مدل‌ها محاسبه و مقادیر شاخص‌های آماری  $R^2$ , RMSE, MAE, MAPE و  $SE$  برای آن‌ها به دست آمد. مقایسه میزان شاخص‌های آماری و نحوه پراکنش داده‌ها در هر سه مدل نشان می‌دهد که نتایج مربوط به مدل‌های ماشینی دارای همبستگی بالاتر و نحوه پراکنش متقارن‌تر و قابل‌قبولی می‌باشد این در حالی است که نتایج مربوط به مدل‌های رگرسیونی دارای ضریب همبستگی پایین و پراکنش نامتقارن‌تری نسبت به روش‌های ماشینی می‌باشد. مقایسه نتایج سه مدل حاکی از برتری مدل GMDH در پیش‌بینی مقادیر رسوب ورودی به آبرگیر حاصل شده است. نتایج شاخص‌های آماری برای سه مدل به ازای الگوی ورودی بهینه مربوط به خود در جدول (۶) ارائه شده است. با توجه به نتایج حاصله می‌توان دید که عملکرد مدل GMDH در مقایسه با سایر مدل‌ها بهتر بوده است، این نتایج با نتایج مطالعات ایوانی و همکاران [۴] و قادری و همکاران [۱۵] مطابقت دارد. مدل‌های رگرسیونی در رتبه‌های دوم، سوم قرار می‌گیرند. نتایج به دست آمده نشان داد روش داده‌کاوی بررسی شده به مراتب نتایج بهتری نسبت به رگرسیونی ارائه می‌کنند

#### توصیه ترویجی

سازه Skimming wall برای کنترل رسوب ورودی به آبرگیر

(GMDH). Journal of Watershed Management Research. 7(13), 218-229. (In ersian)

5. Ghohari, S., Ayyoubzadeh, S.A., Ghodsian, M., and Salehi Neishabouri, A. 2009. The impact of spur dike and submerged vanes on sediment control at lateral intake. Journal of Water and Soil Conservation, 16(2), 35-59. (In Persian)

6. Ivakhnenko, A.G. 1976. The Group Method of Data Handling in Prediction Problems, Soviet Automatic Control of Avtomatika, 9, 21-30.

7. Keshtegar, B., Piri, J., Hussan, W.U., Ikram, K., Yaseen, M., Kisi, O., Adnan, R.M., Adnan, M., and Waseem, M. 2023. Prediction of Sediment Yields Using a Data-Driven Radial M5 Tree Model. Journal of Water 15, 1437. (In Persian)

8. Mehrizi Haeri, A.A. (2013). Data mining: concepts, methods and applications. Master's thesis in economic and social statistics, Faculty of Economics, Allameh Tabatabai University. (In Persian).

9. Mohseni, M. 2022. Prediction of Depth Averaged Velocity in Compound Channels with Vegetated Floodplains Using Gene Expression Programming. Iranian Hydraulic Association Journal of Hydraulics. 17(1), 51-66. (In Persian)

10. Moradi-Nejad, A., Haghiabi, AH., Saneie, M., and Younsei, H. 2018. Laboratory study of the effect of the angle of the skimming wall on the control of sediment in the mouth intake. Journal of Watershed Engineering and Management. 10(1), 28-42. In ersian)

11. Moradi-Nejad, A., Khosrobigi, S., Akbari, M., and Hosseini, SA. 2023. Evaluation of soft calculation methods in estimation of river suspended sediment (Hassanabad station of Tirah River). Soil and water modeling and management. (In Press).

12. Nakato, T., Kennedy, J. F., and Bauerly, D. 1990. Pump station intake shoaling control with submerged vanes. Journal of Hyd. Eng., ASCE, 116(1), 119 - 128 .

13. Neary, V. S., Sotiropoulos, F., and Odgaard, A. J. 1999. Three-dimensional numerical model of lateral-intake inflows. lateral-intake inflows. Journal of Hydraulic Engineering, 25(2), 126-140.

14. Odgaard, J., and Wang, Y. 1991. Sediment management with submerged vanes, ii: applications.

در این مطالعه نسبت رسوب ورودی به آبگیر با استفاده از مدل رگرسیونی چندگانه اینتر و گام به گام و مدل‌های محاسبات نرم، دسته‌بندی گروهی داده‌ها (GMDH) مدل‌سازی و تخمین زده شد. به منظور تخمین پارامتر نسبت رسوب ورودی به آبگیر به عنوان متغیر وابسته و پارامترهای بی‌بعد نسبت دبی ورودی به آبگیر، عدد فرود، زاویه دیوار جداکننده با ساحل و نسبت ارتفاع خارج از بستر بودن دیوار جداکننده به عمق جریان در کانال اصلی به عنوان متغیر مستقل و بی‌بعد به عنوان ورودی مدل‌ها استفاده شد. در تخمین نسبت رسوب ورودی به آبگیر با بهترین عملکرد به ترتیب مربوط به مدل‌های شبکه عصبی دسته‌بندی گروهی داده‌ها (GMDH) با شاخص‌های آماری  $RMSE=0.03$ ,  $MAD=0.039$ ,  $R^2=0.85$  و  $MAPE=0.95$  است. از آنجائی که در طراحی و بهره‌برداری از آبگیرها تخمین نسبت رسوب ورودی به آبگیرها پارامتر مهم و تأثیرگذاری است، می‌توان از روش‌های مبتنی بر داده‌کاوی به عنوان جایگزین روش‌های رگرسیونی برای تخمین بار رسوب ورودی به آبگیر جانبی استفاده کرد. نتایج کلی نشان داد که روش داده‌کاوی نسبت به روش رگرسیونی دقت بالاتری دارند. در نهایت پیشنهاد می‌شود کارایی ورودی‌ها این تحقیق با استفاده از سایر مدل‌های هوشمند (نظیر روش‌های یادگیری عمیق، هیبرید و مدل‌های نروفازی، درخت تصمیم‌گیری و...) و مقایسه نتایج آن‌ها با پژوهش حاضر و نیز کاربرد آن در زاویه‌های مختلف آبگیری دیگر مورد بررسی قرار گیرد.

## منابع

1. Atarzadeh, A., Ayyoubzadeh, S.A., Ghodsian, M., and Salehi Neishabouri, A. 2014. experimental study of the effect of sill, spur dike and submerged vanes on sediment control and bed topography at lateral intakes. Modares Civil Engineering Journal (M.C.E.J). 14(2), 27-38. (In Persian)

2. Barkdoll, B. D., Ettema, R., and Odgaard, J. 1999. Sediment control at lateral diversion: limits and enhancements to vane use. Journal of Hydraulic Engineering, 125(8), 862-870.

3. Biranvand, N., Sepahvand, A., and Haqzadeh, A. 2022. Modeling of suspended sediment using machine learning algorithms in periods of low and high water (case study: Kashkan watershed). Soil and water modeling and management, (3)2. (In Persian)

4. Eivani, Z., Ahmadi, MM., and Qaderi, k. 2016. Estimation of Suspended Sediment Load Concentration in River System using Group Method of Data Handling

19. Sattari, M., Rezazade Jodi, A., Safdari, F., & Kahramanzadeh, F. (2016). Performance evaluation of M5 tree model and support vector regression methods in river suspended sediment modeling. *Journal of Water and Soil Resources Protection*, 6(1), 109-124 (in Persian).
20. Zamani, M., Rabieifar, H., and Rostami, M. 1400. Laboratory and numerical investigation of the effect of the location of the breakwater on the hydraulic and sedimentary conditions of the lateral catchment. *Research paper on watershed management*. 12(24), 287-297. (In Persian)
15. Qaderi, K., Arab, D.R., Teshnelab, M., and Ghazagh, A. 2009. Intelligent Operation Modeling of Reservoirs Using Group Method of Data Handling (GMDH). *Iran Water Resources Research*. 6(3), 55-67. (In Persian)
16. Ramamurthy, A. S., Junying, Q., and Diep, V. 2007. Numerical and experimental study of dividing open-channel flows. *Journal of Hydraulic Engineering*, 133(10), 1135-1144.
17. Sajedi Sabegh, M., and Habibi, M. 2003. Laboratory investigation of using submerged vane and dikes on efficiency of intakes. *Proceeding of the 4th hydraulic conf. shiraz, shiraz University*, 35-42. (In Persian)
18. Salemnia, A., and Shafaei Bajestan, M. 2011. Investigation on the effect of submerged vanes on the amount of sediment entranc of trapezoidal channel into the lateral intake by changing the discharge diversion ratio. 10th Iranian hydraulic conference. Rasht, University of Guilan. (In Persian)



## Abstract

## Using the Regression Method and GMDH Model in Estimating the Ratio of Input Sediment to Lateral Intake

A. Morsdinejad<sup>1</sup>

Received: 2024/06/12 Accepted: 2024/08/23

Due to the changes that occur in the distribution of velocity in the area of the catchment opening, sedimentation usually takes place, which reduces the efficiency of water intake, increases the operational costs for sedimentation operations, and finally changes the direction and contour of the trough. The river drains towards the opposite bank. The use of sediment control structures separating the wall in front of the catchment and at the same time the breakwater in front of the catchment reduces the incoming sediment and increases the catchment efficiency. In the current research, the effect of the separating wall and the breakwater structure in estimating the ratio of incoming sediment to the catchment has been evaluated by laboratory and data mining and multiple regression methods. First, by performing dimensional analysis, dimensionless ratios were extracted and the relationship between the variables and their value in the experiments was determined. Using XLSTAT and SPSS statistical software, equations for the relationship between independent and dependent variables were extracted from the step-by-step and inter method. After obtaining the equations, the relative error of each equation was calculated. Then the best equation with high  $R^2$  and low relative error was selected and proposed. In the next step, modeling was done with the methods of artificial neural networks and the method of group data control (GMDH) and the best method was selected to estimate the ratio of input sediment to the catchment. The results showed that in estimating the ratio of input sediment to the catchment, the best performance is related to the model (GMDH) with statistical indices,  $R^2=0.850$ ,  $MAD=0.03$ ,  $RMSE=0.039$  and  $MAPE=26.95$ . In the step-by-step regression method,  $R^2=0.38$ ,  $RMSE=4.99$  and inter regression method,  $R^2=0.76$ ,  $RMSE=4.16$ . also data-mining methods compared to the method Regression has higher accuracy.

**Keywords:** Aquifer, Breakwater, Separating wall, Sediment, Data-mining.

1. Assist. Prof., Soil Conservation and Watershed Management Research Department, Markazi Agricultural and Natural Resources Research and Education Center, Arak, Agricultural Research Education & Extension Organization (AREEO). Arak, Iran, Email: amir\_24619@yahoo.com