

## تخمین ضریب پخش عرضی در انتقال آلودگی در رودخانه‌های عریض با استفاده از محاسبات تکاملی

اسما براهویی نژاد<sup>۱</sup>، علیرضا قائمی<sup>۲</sup>، سید آرمان هاشمی منفرد<sup>۳\*</sup>، غلامرضا عزیزیان<sup>۳</sup>، محسن دهقانی درمیان<sup>۲</sup>  
 ۱. دانشجوی کارشناسی ارشد عمران آب و سازه‌های هیدرولیکی، دانشکده مهندسی شهید نیکبخت، دانشگاه سیستان و

بلوچستان، زاهدان

۲. دانشجوی دکتری مهندسی عمران، دانشکده مهندسی شهید نیکبخت، دانشگاه سیستان و بلوچستان، زاهدان

۳. دانشیار گروه مهندسی عمران، دانشکده مهندسی شهید نیکبخت، دانشگاه سیستان و بلوچستان، زاهدان

(تاریخ دریافت ۱۳۹۸/۰۲/۱۰؛ تاریخ تصویب ۱۳۹۸/۰۴/۲۴)

### چکیده

در قرن اخیر با رشد روزافزون جمعیت شهرنشین، مشکلات متعددی در زمینه آلودگی و کیفیت منابع آبی مطرح شده است. بنابراین، شناخت و مطالعه فرایندهای اختلاط و انتقال مواد در رودخانه‌ها از جمله فعالیت‌های مهم در برنامه‌های مدیریت منابع آب به شمار می‌آید. بین فرایندهای اختلاط، پس از پدیده انتشار طولی، فرایند انتشار عرضی آلودگی، تأثیرگذارترین پارامتر محسوب می‌شود. با توجه به اهمیت انتقال و چگونگی انتشار آلودگی در رودخانه‌ها، تخمین ضریب پخش عرضی انتقال آلودگی در جریان‌های سطحی با استفاده از ماشین بردار پشتیبان (SVM) و با بهره‌گیری از دو کرنل تابع پایه شعاعی و چندجمله‌ای و مدل درخت (MT)، هدف اصلی پژوهش حاضر است. برای تخمین ضریب پخش عرضی از ۱۸۷ سری داده که شامل عمق جریان، سرعت جریان، سرعت برشی و عرض کانال می‌شود، استفاده شده است. نتایج به‌دست‌آمده از معیارهای ارزیابی نشان داد مدل SVM-Poly دقت بیشتری ( $R=0/992$  و  $OI=0/982$ ) نسبت به مدل SVM-RBF دارد. همچنین مدل SVM-RBF دقت بیشتری ( $R=0/968$  و  $OI=0/950$ ) نسبت به مدل MT ( $R=0/966$  و  $OI=0/946$ ) در تخمین این پارامتر در مرحله آموزش داشت. مقادیر به دست‌آمده از  $D_T$  در مرحله تست هم ارزیابی شدند و مشخص شد که SVM-RBF با داشتن کمترین خطا ( $RMSE=0/029$ ) توانایی بهتری در تخمین  $D_T$  دارد. علاوه بر این، مقایسه عملکرد روش‌های هوشمند با روابط تجربی بیان می‌کند که روابط تجربی دقت قابل قبولی نداشته‌اند.

**کلیدواژه‌گان:** انتقال آلودگی، رودخانه، ضریب پخش عرضی، ماشین بردار پشتیبان، مدل درخت.

## مقدمه

جریان و سرعت برشی هستند. رادرفورد با جمع‌آوری و بررسی یک سری از داده‌های میدانی متوجه شد که مقدار بی‌بعد  $\frac{D_T}{HU_*}$  در بازه ۰/۱ تا ۰/۲۶ متغیر است [۵]. دنگ و همکارانش در سال ۲۰۰۱ با بررسی سری داده‌های بی‌بعد جمع‌آوری‌شده توسط رادرفورد [۵]، رابطه ۲ را برای تخمین ضریب پخش عرضی در کانال‌های بدون انحنا ارائه دادند [۶].

$$D_T = 0.145 HU_* \quad (2)$$

همچنین، بنسالی یک معادله تجربی با استفاده از پارامترهای ورودی عرض کانال (W)، عمق جریان (H) و سرعت برشی ( $U_*$ ) برای برآورد این فاکتور از ضریب پخش ارائه داد که به صورت رابطه ۳ بیان شده است [۷].

$$D_T = HU_* \left[ 0.002 \left( \frac{W}{H} \right)^{1.5} \right] \quad (3)$$

از طرفی، با توجه به پیشرفت علم در زمینه‌های مختلف و کاربرد فراوان انواع روش‌های هوش مصنوعی (AI) مانند شبکه‌های عصبی مصنوعی (ANNs)، سیستم تطبیقی استنتاج عصبی فازی (ANFIS)، برنامه بیان ژن (GEP)، مدل درخت (MT) و ماشین بردار پشتیبان (SVM)، راه‌کارهای مؤثری را برای حل مسائل گوناگون در زمینه مهندسی محیط زیست ارائه داده‌اند [۸-۱۰]. همچنین، کاربرد وسیع تکنیک‌های مختلف هوش مصنوعی از جمله رگرسیون چندجمله‌ای تکاملی (EPR)، الگوریتم مدل درخت (MT) و روش دسته‌بندی گروهی داده‌ها بر پایه منطق فازی (NF-GMDH) در ضریب پخش طولی هم‌انکارندگی است [۱۱-۱۴]. به طور مثال، ساهای با استفاده از شبکه عصبی مصنوعی به تخمین ضریب پخش طولی پرداخت. او پارامترهای ورودی را عمق جریان، عرض کانال، سرعت جریان، تنش برشی بستر و انحراف کانال انتخاب کرد. نتایج پژوهش ایشان نشان داد ANN دقت بیشتری نسبت به روش‌های تجربی دارد

مطالعه روی کیفیت آب‌های سطحی اهمیت زیادی دارد. با توجه به اینکه آب‌های سطحی یکی از منابع عمده تأمین آب آشامیدنی انسان‌ها، آب لازم برای کشاورزی و صنایع از رودخانه‌هاست، پژوهش درباره آن ضرورت دارد. متأسفانه، گاهی فاضلاب صنعتی و پساب در رودخانه‌ها تخلیه می‌شود. اگر مکانیسم انتقال و پخش آلودگی در رودخانه‌ها با ژئومتری‌های مختلف مشخص باشد، می‌توان برای کاهش آثار آلودگی بر سلامت عموم جامعه انسان‌ها با طرح مسئله اختلاط و تقویت قدرت خودپالایندگی رودخانه‌ها برنامه‌ریزی کرد [۱].

فرایند انتشار آلودگی در رودخانه‌ها به سه مرحله اصلی تقسیم می‌شود: ابتدا اختلاط اولیه آغاز می‌شود که هدف آن، تعیین ضریب پخش قائم است. آلودگی پخش می‌شود و تا جایی که پخش غلظت در عمق یکنواخت شود، ادامه می‌یابد. پس از اختلاط اولیه، اختلاط کامل شروع می‌شود و تا مکانی که آلودگی در مقطع عرضی منتشر می‌شود، ادامه می‌یابد. تخمین ضریب پخش عرضی و طول اختلاط، از اهداف مطالعات ردیابی در این مرحله است. در نهایت، منطقه دوردست آغاز خواهد شد که می‌توان آن را حد فاصل اختلاط کامل و جایی که غلظت آلودگی قابل تشخیص نباشد، در نظر گرفت. علاوه بر انتقال ذرات توسط جابه‌جایی، پدیده غالبی که در این ناحیه وجود دارد، انتشار طولی است که سبب منبسط شدن ابر آلودگی در راستای طولی (در جهت سرعت طولی) می‌شود. با توجه به موارد یادشده، فاکتور ضریب پخش که به عنوان پارامتر بسیار مهمی در تعیین چگونگی پخش شدن آلودگی، کارکرد فراوانی دارد، از سه جزء قائم، عرضی و طولی تشکیل شده است و تخمین دقیق آن‌ها در مهندسی رودخانه و محیط زیست اهمیت زیادی دارد [۱-۳].

مطالعات اندکی برای تخمین ضریب پخش عرضی انجام شده است. به طور مثال، فیشر و همکارانش براساس مطالعات آزمایشگاهی در کانال‌های مستقیم (بدون انحنا و خم)، رابطه‌ای تجربی را برای تخمین ضریب اختلاط عرضی ارائه دادند که به صورت رابطه ۱ بیان می‌شود [۴]:

$$D_T = 0.15 HU_* \quad (1)$$

که در رابطه یادشده، H و  $U_*$  به ترتیب بیان‌کننده عمق

1. Artificial Intelligence
2. Artificial Neural Networks
3. Adaptive Neuro-fuzzy Inference System
4. Gene-expression programming
5. Model Tree
6. Support Vector Machine
7. Evolutionary Polynomial Regression
8. Group Method of Data Handling

رودخانه‌ها دقت قابل قبولی داشته باشد. بنابراین، با توجه به اهمیت فاکتور ضریب پخش عرضی در رودخانه‌ها و آنچه پیش‌تر از توانایی روش‌های مختلف هوش مصنوعی در حل مسائل پیچیده گفته شد، تخمین ضریب پخش عرضی با استفاده از دو روش نوین هوشمند شامل مدل درخت و ماشین بردار پشتیبان با استفاده از دو کرنل چندجمله‌ای و تابع پایه شعاعی، به عنوان هدف اصلی تحقیق حاضر قرار گرفته است. از این رو، از ۱۸۷ سری داده شامل عمق جریان، عرض کانال، سرعت جریان و سرعت برشی به عنوان پارامترهای ورودی و ضریب پخش عرضی به عنوان پارامتر هدف، که از مطالعات گذشته جمع‌آوری شده، استفاده شده است.

### مواد و روش‌ها

#### روش تحقیق

#### ماشین بردار پشتیبان (SVM)

ماشین بردار پشتیبان از جمله روش‌های یادگیری با ناظر است که برای طبقه‌بندی و رگرسیون به کار برده می‌شود و در سال ۱۹۹۲ توسط وپنیک، ریاضیدان روسی، و چرونکیس بر پایه تئوری یادگیری آماری ارائه شد. این روش از یک ابرصفحه خطی تفکیک‌گر بهینه برای دسته‌بندی داده‌ها استفاده می‌کند که این ابرصفحه از حداکثر کردن حاشیه بهره می‌برد. به همین دلیل باید یک مسئله بهینه‌سازی غیر خطی مقید حل شود. از جمله دسته‌بندی‌کننده‌هایی که بدون دانستن مدل توزیع داده‌ها، دسته‌بندی را انجام می‌دهند و نیازی به دانستن ساختار کلی مسئله ندارند. این روش بردارهای ورودی را با استفاده از یک تابع کرنل به فضای بزرگ‌تری می‌برد تا بتواند تفکیک‌پذیری خطی را بهتر انجام دهد [۱۶]. در یک مدل رگرسیونی SVM لازم است که وابستگی تابعی متغیر وابسته  $y$  به مجموعه‌ای از متغیرهای مستقل  $x$  تخمین زده شود. مانند دیگر مسائل رگرسیونی، فرض بر این است که رابطه بین متغیرهای وابسته و مستقل توسط یک تابع معین  $f(x)$  به علاوه یک مقدار اضافی نویز مشخص شود [۸]:

$$y = f(x) + \text{noise} \quad (۵)$$

بنابراین، پیدا کردن فرم تابع  $f(x)$  اهمیت زیادی دارد

[۱۱]. اعتمادشهودی و تقی‌پور عملکرد مدل درخت را برای تخمین ضریب پخش آلودگی ارزیابی کردند. روابط به‌دست‌آمده از روش آنها توانست مقادیر ضریب پخش را با دقت قابل قبولی پیش‌بینی کند. همچنین، ارزیابی نتایج پژوهش آنها با نتایج مطالعات گذشته نشان داد روش پیشنهادی خطای کمتری دارد [۱۲]. نجف‌زاده و تفرج نوروز با استفاده از روش دسته‌بندی گروهی داده‌ها بر پایه الگوریتم بهینه‌سازی جمعیت ذرات (PSO) ضریب پخش آلودگی در رودخانه‌ها را پیش‌بینی کردند. آنها پارامترهای مؤثر را عرض کانال، عمق جریان، سرعت جریان و سرعت برشی بستر، در نظر گرفتند و نتایج مدل پیشنهادی را با روابط تجربی مقایسه کردند که در آن، مدل پیشنهادی آنها عملکرد قابل قبول‌تری داشت [۱۳].

رضایی و همکارانش از روش رگرسیون چندجمله‌ای تکاملی به منظور تخمین ضریب پخش طولی استفاده کردند. براساس نتایج پژوهش آنها، روش EPR توانایی خوبی نسبت به روابط تجربی و روش‌هایی مانند MT و برنامه‌نویسی ژنتیک (GP) در تخمین ضریب پخش طولی داشت. همچنین، آنها عرض کانال را به عنوان تأثیرگذارترین پارامتر روی ضریب پخش طولی معرفی کردند [۱۴].

با توجه به بررسی‌های متعدد انجام‌شده، مطالعات اندکی در زمینه ضریب پخش عرضی در رودخانه‌های عریض انجام شده است. هوآی و همکاران (۲۰۱۸) از برنامه‌نویسی ژنتیک (GP) به منظور بررسی عملکرد این روش در تخمین اختلاط عرضی در کانال‌ها، استفاده کردند که به ایجاد شدن رابطه ۴ منجر شد [۱۵].

$$D_T = HU_* \left( \frac{0.121 \left( \frac{B}{H} \right)}{\left( \frac{B}{H} \right) + 0.222 \left( \frac{U}{U_*} \right) - 1.99} \right) \quad (۴)$$

از آنجا که ممکن است روابط تجربی متعددی که متناسب با رودخانه خاصی ارائه شده است، برای سایر رودخانه‌ها دقت کافی نداشته باشند، محققان تصمیم گرفتند با جمع‌آوری داده‌های استفاده‌شده از سایر رودخانه‌ها، مدل‌هایی ارائه بدهند که برای تمامی

درخت فرعی با یک برگ صورت می‌گیرد. بنابراین، مرحله دوم در طراحی مدل درختی، شامل هرس کردن درخت رشدیافته و جایگزینی درختان فرعی با توابع رگرسیونی چندمتغیره خطی است. با این تکنیک، فضای پارامترهای ورودی به زیرفضاهای کوچک‌تر تقسیم شده و در هر یک از آنها، نوعی مدل رگرسیونی خطی برازش داده می‌شود. بعد از آنکه مدل خطی به دست آمد، برای کمینه کردن خطای تخمین با حذف کردن پارامترها، ساده‌سازی مدل انجام می‌شود. توضیحات بیشتر در مورد مدل درخت در منابع ارائه شده است [۱۹ و ۲۰].

#### داده‌های استفاده شده

در مطالعه حاضر، ۱۸۷ سری داده آزمایشگاهی شامل عمق جریان (H)، عرض رودخانه (W)، سرعت جریان (U) و سرعت برشی (U\*) که از مطالعات فیشر (۱۹۶۷)، رادفورد (۱۹۶۷)، بنسال (۱۹۷۱)، بلتاس و دی (۱۹۷۴)، انگمن و کلرهاس (۱۹۷۴)، یاسوکورا و سایره (۱۹۷۶)، وبل و شاتزمن (۱۹۷۸)، دنگ و همکاران (۲۰۰۱)، لانگ و همکاران (۲۰۰۲)، جئون و همکاران (۲۰۰۷)، طباطبایی و همکاران (۲۰۱۳)، پیلچی و همکاران (۲۰۱۶) و هوای و همکاران (۲۰۱۸) جمع‌آوری شده است، برای تخمین ضریب پخش عرضی (مطابق رابطه ۷) استفاده شده است:

$$D_T = f(W, H, U, U_*) \quad (7)$$

همچنین، برای مدل‌سازی روش‌های هوشمند مصنوعی، ۷۵ درصد از کل داده‌ها (۱۴۰ سری داده) به انجام مرحله آموزش اختصاص می‌شود و یک چهارم باقی‌مانده (۴۷ سری داده) نیز برای صحت‌سنجی یا آزمایش در نظر گرفته می‌شود. خصوصیات آماری داده‌های پارامترهای استفاده شده برای مدل‌سازی در تحقیق حاضر در جدول ۱ ارائه شده است.

#### ساختار روش MT برای تخمین $D_T$

مدل درختی پارامتر خروجی محاسباتی را به صورت مجموعه قوانین اگر و آن‌گاه (معادلات خطی) و با توجه به محدوده پارامترهای ورودی نتیجه می‌دهد. بنابراین، بعد از تفکیک داده‌ها به دو بخش آموزش و آزمایش، مدل درختی M5 روی داده‌های آموزش برازش داده می‌شود. در این مرحله، داده‌ها به گروه‌هایی دسته‌بندی شده‌اند و برای

که بتواند به صورت صحیح، مواردی که SVM تجربه نکرده است را پیش‌بینی کند. این تابع به وسیله آموزش مدل SVM روی مجموعه داده‌های آموزش که شامل فرایند بهینه‌سازی دائمی تابع خطاست، در دسترسی قرار می‌گیرد [۱۷].

#### مدل درخت (MT)

برای نخستین بار کوینلان [۱۸] مدل درخت تصمیم موسوم به M5 را برای پیش‌بینی داده‌های پیوسته ارائه کرد. این مدل، به خلاف مدل‌های درخت تصمیم معمول که کلاس یا رده‌های گسسته را به عنوان خروجی ارائه می‌کنند، نوعی مدل خطی چندمتغیره را برای داده‌ها در هر گره از مدل درختی می‌سازد. مدل‌های درخت تصمیم‌گیری از دو مرحله ایجاد درخت و هرس کردن آن، ایجاد می‌شود. به منظور تولید یک درخت تصمیم، از نوعی الگوریتم استنتاجی یا معیار تقسیم استفاده می‌شود. در مدل‌سازی M5، انحراف معیار کاهشی به عنوان معیاری برای ارزیابی دقت هر مدل خطی در هر گره به کار برده می‌شود و کاهش مورد انتظار در این خطا را به عنوان نتیجه آزمون هر ویژگی (صفت) در آن گره در نظر می‌گیرند. کاهش انحراف معیار (SDR)<sup>۱</sup> از رابطه ۶ به دست می‌آید [۹]:

$$SDR = sd(T) - \sum \frac{|T_i|}{|T|} sd(T_i) \quad (6)$$

که T بیان‌کننده یک سری نمونه‌های قبل از تقسیم‌بندی،  $T_i$  نشان‌دهنده داده‌ای است که بعد از تقسیم‌بندی حاصل می‌شود و با توجه به پارامتر تقسیم انتخاب شده به یک زیرفضا می‌افتد و sd بیان‌کننده انحراف معیار است. به دلیل فرایند انشعاب، داده‌های قرارگرفته در گره‌های فرزند، انحراف معیار کمتری نسبت به گره مادر دارند و بنابراین خالص‌ترند. پس از حداکثرسازی تمامی انشعاب‌های ممکن، M5 ویژگی را انتخاب می‌کند که کاهش مورد انتظار را بیشینه کند. در این نوع تقسیم‌بندی (کلاسه‌بندی) خروجی مدل M5 به صورت ساختار شبیه درخت می‌شود که بیش‌برازش را در پی دارد. برای جلوگیری از این فرایند، مدل M5 وارد مرحله هرس کردن درخت ایجادشده، می‌شود. این کار با جایگزینی یک

ساختار روش SVM برای تخمین شاخص  $D_T$  به منظور تخمین ضریب پخش طولی و عرضی، ماشین بردار پشتیبان با استفاده از کرنل‌های تابع پایه شعاعی و چندجمله‌ای ارزیابی شده است. برای بهینه کردن پارامترها از جست‌وجوی شبکه‌ای میان پارامترهای ورودی در مرحله آموزش استفاده شده است. پارامترهای بهینه در تابع پایه شعاعی (گاما و سیگما) و در چندجمله‌ای (گاما، سیگما و درجه) مقادیر بهینه‌ای هستند که باید تعیین بشوند که در جدول ۳ نشان داده شده‌اند.

هر دسته یک مدل رگرسیون چندمتغیره خطی ارائه شده است. نتایج مدل پیشنهادی درختی نشان داد با وجود چهار متغیر ورودی، هفت معادله رگرسیونی خطی به صورت قوانین اگر و آن‌گاه برای تخمین شاخص  $D_T$  تولید شدند. در تحقیق حاضر، تمام داده‌های ورودی که به هفت قسمت تقسیم شده‌اند در هفت معادله خطی چندمتغیره صدق می‌کنند که پارامتر خروجی به دست آمده از هفت معادله بیان‌کننده  $D_T$  محاسباتی است. در جدول ۲ نمونه‌ای از قوانین تولیدشده توسط M5 ارائه شده است.

جدول ۱. ویژگی‌های آماری پارامترهای استفاده‌شده در مطالعه حاضر

پارامتر	واحد	کمترین مقدار	بیشترین مقدار	انحراف معیار	میانگین
W	متر (m)	۰/۳	۶۳۸	۱۰۷/۶۱۸	۳۹/۷۴
H	متر (m)	۰/۰۱۲	۱۹/۱۴	۲/۴۴۴	۰/۷۸۴
U	متر بر ثانیه (m/s)	۰/۰۲۳	۱/۹۸	۰/۳۷۲	۰/۳۶۴
U*	متر بر ثانیه (m/s)	۰/۱۶۳	۰/۰۰۳۵	۰/۰۲۹	۰/۰۳۱
$D_T$	مجذور متر بر ثانیه ( $m^2/s$ )	۰/۰۰	۱/۱۵۶	۰/۱۴۵	۰/۰۳۸

جدول ۲. سه قانون نخست به دست آمده از اجرای M5 برای تخمین شاخص  $D_T$

شماره قانون	اگر (IF)	آن‌گاه (THEN)
قانون اول	$H \leq 0/328$	$D_T = + 0/0055 \times H + 0/0169 \times U - 0/0043$
قانون دوم	$H \leq 3/42$ $U \leq 0/905$ $U > 0/355$	$D_T = + 0/0441 \times H + 0/1528 \times U - 0/1244$
قانون سوم	$H \leq 3/95$ $H \leq 2/84$ $B > 86/95$	$D_T = 0/0776 \times H - 0/277$

جدول ۳. مقادیر بهینه مربوط به ماشین بردار پشتیبان

	$\gamma$	$\sigma$	t
RBF	۱۸/۰۴۶	۱۷/۴۶۷	-----
Poly	۰/۰۹۰	۱۶/۵۹۱	۳

میانگین مطلق خطا (MAE<sup>۲</sup>) و شاخص کلی عملکرد مدل (OI<sup>۴</sup>)، هستند که در روابط ۸-۱۱ نشان داده شده‌اند [۲۸ و ۲۹]:

مقایسه عملکرد مدل‌های پیشنهادی در این قسمت، به منظور مقایسه عملکرد مدل‌های ارائه‌شده در مراحل آموزش و تست، از شاخص‌های آماری استفاده شده است. این شاخص‌ها شامل ضریب همبستگی (R<sup>۱</sup>)، ریشه میانگین مربعات خطا (RMSE<sup>۳</sup>)،

2. Root Mean Square Error  
3. Mean Absolute Error  
4. Overall Index

1. Correlation Coefficient

جدول ۴. ارزیابی عملکرد مدل‌های پیشنهادی در مرحله آموزش

مدل	R	RMSE	MAE	OI
SVM-Poly	۰/۹۹۲	۰/۰۲۰	۰/۰۱۰۵	۰/۹۸۲
SVM-RBF	۰/۹۶۸	۰/۰۳۹	۰/۰۱۰۵	۰/۹۵۰
MT	۰/۹۶۶	۰/۰۴۱	۰/۰۱۵	۰/۹۴۶

در ادامه، شکل ۱ مقادیر ضریب پخش عرضی پیش‌بینی شده توسط مدل‌های پیشنهادی در مقابل مقادیر مشاهداتی آنها را نشان می‌دهد. در یک مقایسه کیفی، بیشتر مقادیر  $D_T$  در شکل ۱ در بازه صفر تا ۰/۲ تمرکز بیشتری دارند و با توجه به محدوده خطای ۳۰ درصد می‌توان گفت که خطای بیشتر مقادیر پیش‌بینی شده توسط مدل‌ها، حدود ۳۰ درصد است. همین‌طور مشخص است که نقاط مربوط به روش SVM-Poly به نیمساز ربع اول و سوم نزدیک‌تر است. علاوه بر این، می‌توان گفت که بیشتر نقاط پیش‌بینی شده  $D_T$  توسط مدل‌های SVM و MT از مقادیر واقعی آنها کوچک‌ترند.

پارامترهای آماری برای صحت‌سنجی و کارایی مدل‌های ماشین بردار پشتیبان و مدل درخت در مرحله تست هم استفاده شد و نتایج آن در جدول ۵ آورده شده است.

همان‌طور که جدول ۵ نشان می‌دهد، به رغم اینکه SVM-Poly ضریب همبستگی (۰/۹۷۲) بیشتری نسبت به SVM-RBF (۰/۹۷۱) و MT (۰/۹۳۴) دارد، مقادیر شاخص‌های آماری دیگر مانند ریشه میانگین مربعات خطا دلالت بر ضعف مدل SVM-Poly است و می‌توان نتیجه گرفت که یک پارامتر به تنهایی نمی‌تواند تعیین‌کننده باشد. مقدار شاخص آماری RMSE و OI برای مدل پیشنهادی SVM-RBF در این مرحله به ترتیب ۰/۰۲۹ و ۰/۹۴۳ به دست آمده است که نشان می‌دهند مدل SVM-RBF عملکرد خوب و دقت زیادی در مرحله تست داشته است. همچنین، به رغم اینکه مدل SVM-Poly در مقایسه با روش‌های SVM-RBF و MT در مرحله آموزش کمترین خطا (RMSE=۰/۰۲۰ و MAE=۰/۰۱۰۵) در پیش‌بینی ضریب پخش عرضی را داشت، در مرحله تست با داشتن خطای نسبتاً بیشتر (RMSE=۰/۰۷۶ و MAE=۰/۰۲۷) نسبت به مدل‌های SVM-RBF (RMSE=۰/۰۲۹ و MAE=۰/۰۱۱) و MT (RMSE=۰/۰۳۹ و MAE=۰/۰۱۵) نتوانست عملکرد خوبی از خود نشان دهد.

$$R = \frac{\sum_{i=1}^N (D_{TPre}^i - D_{TPre}^{mean})(D_{TObs}^i - D_{TObs}^{mean})}{\sqrt{\sum_{i=1}^N (D_{TObs}^i - D_{TObs}^{mean})^2 \sum_{i=1}^N (D_{TPre}^i - D_{TPre}^{mean})^2}} \quad (8)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (D_{TObs}^i - D_{TPre}^i)^2}{N}} \quad (9)$$

$$MAE = \frac{\sum_{i=1}^N |D_{TPre}^i - D_{TObs}^i|}{N} \quad (10)$$

$$OI = \frac{1}{2} \left( 2 - \frac{RMSE}{D_{TObs}^{max} - D_{TObs}^{min}} - \frac{\sum_{i=1}^N (D_{TObs}^i - D_{TPre}^i)^2}{\sum_{i=1}^N (D_{TObs}^i - D_{TObs}^{mean})^2} \right) \quad (11)$$

که  $D_{TObs}^i$  مقادیر مشاهده‌ای،  $D_{TPre}^i$  مقادیر پیش‌بینی شده،  $D_{TPre}^{mean}$  میانگین داده‌های مشاهده‌ای،  $D_{TObs}^{mean}$  میانگین داده‌های پیش‌بینی و N تعداد داده‌هاست.

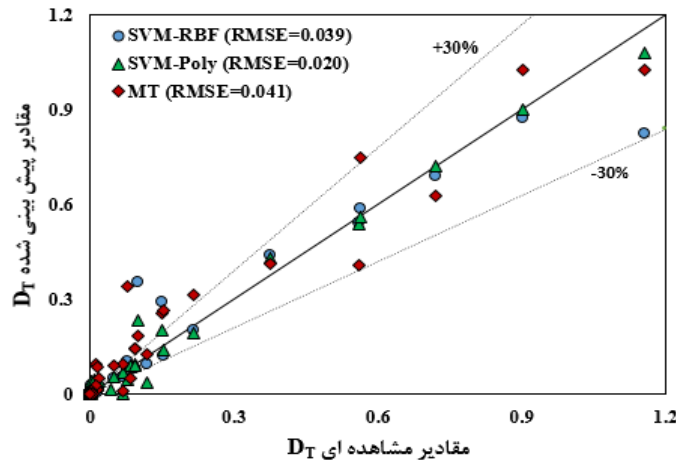
پارامترهای R، OI، RMSE و MAE از جمله معیارهای ارزیابی‌اند. مقادیر RMSE و MAE از جنس پارامتر خروجی هستند و مقادیر بین صفر تا عدد مثبت بزرگ را اختیار می‌کنند. این مقادیر هرچه به صفر نزدیک‌تر باشند، نشان‌دهنده خطای کمتر مدل است. ضریب همبستگی در بازه -۱ و ۱ تغییر می‌کند که هرچه به یک نزدیک‌تر باشد، مدل دقت زیادی دارد. مقادیر پارامتر OI هم بین صفر و یک هستند که هرچه این مقدار به یک نزدیک‌تر باشد، نشان‌دهنده عملکرد بهتر مدل است.

### نتایج و بحث

ارزیابی عملکرد مدل‌های پیشنهادی در تخمین ضریب پخش عرضی در دو مرحله آموزش و تست، به ترتیب در جدول‌های ۴ و ۵ آورده شده است. با توجه به جدول ۴، در تخمین ضریب پخش عرضی ( $D_T$ ) می‌توان گفت که SVM که از کرنل چندجمله‌ای استفاده می‌کند توانست عملکرد قابل قبول‌تری را در مقایسه با سایر روش‌های پیشنهادی داشته باشد. شاخص‌های آماری R (۰/۹۹۲)، RMSE (۰/۰۲۰) و OI (۰/۹۸۲) بیان‌کننده این موضوع هستند. علاوه بر این، بیشترین مقدار RMSE که برابر ۰/۰۴۱ است، مربوط به مدل درخت می‌شود که توانست مقادیر  $D_T$  را با بیشترین خطا نسبت به روش ماشین بردار پشتیبان، تخمین بزند.

را نسبت به رابطه ارائه شده توسط فیشر را داشت. برای درک بهتر مقادیر پارامتر خطای RMSE و MAE مربوط به هر مدل در مرحله آموزش توسط شکل ۲ نشان داده شده است. با توجه به شکل مشخص است که مدل SVM-RBF و رابطه ارائه شده توسط بنسال به ترتیب کمترین و بیشترین مقادیر RMSE و MAE را دارند.

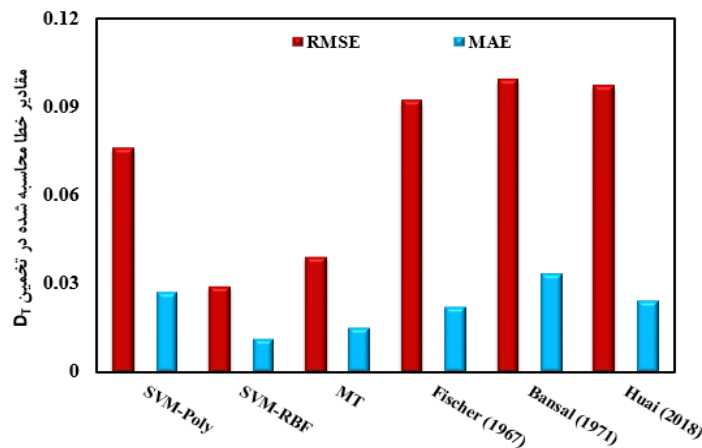
دقت روابط تجربی ارائه شده با یکدیگر در مرحله تست نیز ارزیابی شدند. نتایج نشان داد رابطه فیشر [۱] با داشتن  $R=0/95$  و  $RMSE=0/92$  در مقایسه با روابط ۳ و ۴ دقت بهتری در تخمین  $D_T$  داشت. اگرچه در مقایسه با روش‌های نوین هوش مصنوعی، مدل SVM-Poly که کمترین دقت را در بین روش‌های SVM-RBF و MT داشت، خطای کمتری



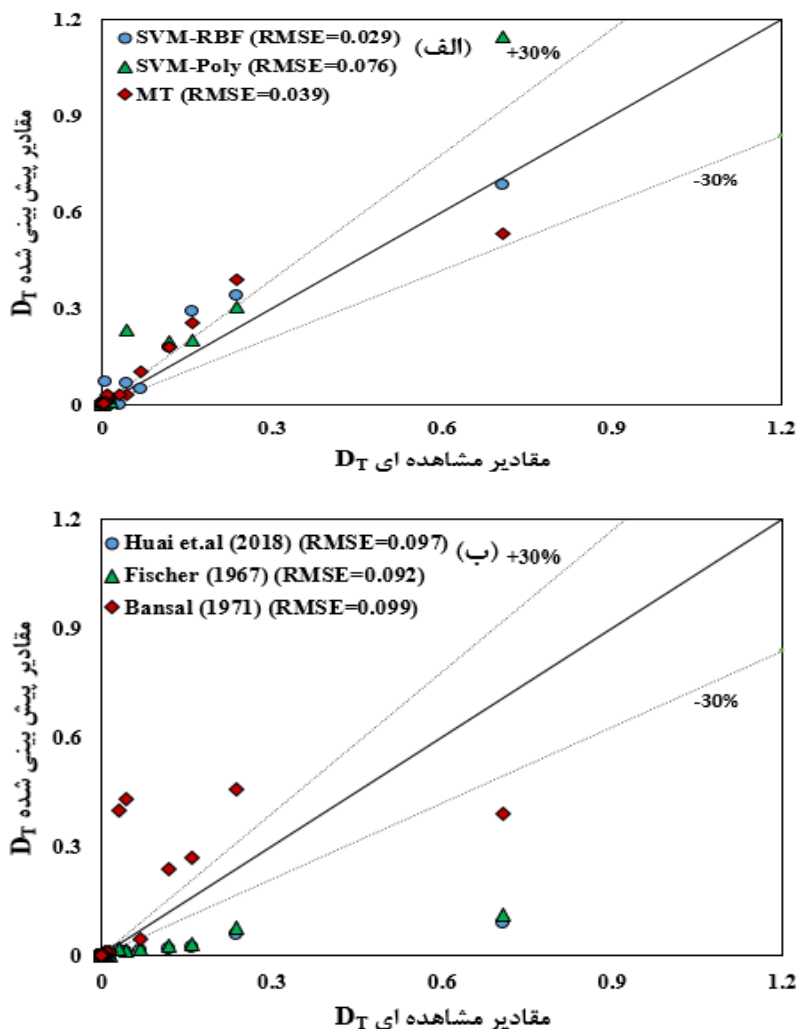
شکل ۱. مقادیر مشاهداتی در برابر مقادیر پیش‌بینی شده  $D_T$  در مرحله آموزش

جدول ۵. ارزیابی مدل‌های پیشنهادی و روابط تجربی برای تخمین  $D_T$  در مرحله تست

مدل	R	RMSE	MAE	OI
SVM-Poly	۰/۹۷۲	۰/۰۷۶	۰/۰۲۷	۰/۷۰۱
SVM-RBF	۰/۹۷۱	۰/۰۲۹	۰/۰۱۱	۰/۹۴۳
MT	۰/۹۳۴	۰/۰۳۹	۰/۰۱۵	۰/۹۰۸
Fischer (۱۹۶۷)	۰/۹۵	۰/۰۹۲	۰/۰۲۲	۰/۵۷۳
Bansal (۱۹۷۱)	۰/۶۶	۰/۰۹۹	۰/۰۳۳	۰/۵۱۷
Huai et al. (۲۰۱۸)	۰/۹۵	۰/۰۹۷	۰/۰۲۴	۰/۵۳۶



شکل ۲. مقادیر محاسباتی پارامتر RMSE و MAE برای هر مدل در مرحله تست



شکل ۳. مقادیر مشاهداتی در برابر مقادیر پیش‌بینی شده  $D_T$  در مرحله تست الف) روش‌های هوشمند، ب) روابط تجربی

می‌گیرد. حال R بیشتر و هم‌پس‌طور RMSE و MAE کمتر، سبب کاهش مقدار تابع هدف شده و بیان‌کننده زیاد بودن دقت مدل مد نظر است. مقدار تابع هدف از رابطه ۱۲ به دست می‌آید [۲۹]:

$$OBJ = \left( \frac{NO_{DT_{tr}} - NO_{DT_{te}}}{NO_{DT_{total}}} \right) \frac{MAE_{DT_{tr}}}{R^2_{DT_{tr}}} + \frac{2NO_{DT_{te}} \times MAE_{DT_{te}}}{NO_{DT_{total}} \times R^2_{DT_{te}}} \quad (12)$$

در این رابطه  $NO_{validation}$  و  $NO_{learning}$  به ترتیب تعداد داده‌های مربوط به مرحله آموزش و تست است و هم‌پس‌طور MAE پارامتر متوسط قدر مطلق خطاست. هم‌پس‌طور، MAE مقدار میانگین مثبت خطاها در یک مجموعه از مقادیر پیش‌بینی شده را محاسبه می‌کند و واحد

همچنین، شکل ۳ مقادیر پیش‌بینی شده  $D_T$  توسط مدل‌های پیشنهادی در مقابل مقادیر مشاهداتی آنها در مرحله تست را نشان می‌دهد. در یک مقایسه کیفی، بیشتر مقادیر  $D_T$  در شکل ۳ همانند شکل ۱ در بازه صفر تا ۰/۲ تمرکز بیشتری دارند و همچنین می‌توان گفت که همگرایی مقادیر  $D_T$  پیش‌بینی شده توسط مدل SVM-RBF به خط ایده‌آل در مقایسه با مدل SVM-Poly و MT در مرحله تست بیشتر است و بیشتر نقاط پیش‌بینی شده توسط این مدل در محدوده خطای ۳۰ درصد هستند.

از نظر دیگر باید گفت که محاسبه هر پارامتر خطا به صورت جداگانه، نمی‌تواند به تنهایی نشان‌دهنده دقت هر مدل باشد. از این رو، از پارامتر آماری دیگری به نام تابع هدف (Objective function) استفاده شده است که هم‌زمان پارامترهای R، RMSE و MAE را در نظر



می‌شود. به بیان دیگر، روشی برای ایجاد تغییر در ورودی‌های یک مدل آماری به صورت سازمان یافته (سیستماتیک) است که بتوان تأثیرات این تغییرها را در خروجی مدل پیش‌بینی کرد. علت اهمیت بحث تحلیل حساسیت در این است که در دنیای واقعی پارامترهای مدل (ضرایب تابع هدف، مقادیر سمت راست، تعداد متغیرها، تعداد محدودیت‌ها) تغییر می‌کند و این تغییرات مربوط به شرایط محیطی است [۲۹]. در تحقیق حاضر، به منظور مشخص کردن تأثیر هر یک از پارامترهای ورودی به مدل پیشنهادی برای تخمین ضریب پخش عرضی، آنالیز حساسیت روی مدل SVM-RBF انجام می‌شود. برای انجام آنالیز حساسیت، در هر نوبت یکی از پارامترهای ورودی شبکه را حذف می‌کند و با داشتن سه ورودی و یک خروجی، مدل‌سازی ضریب پخش عرضی انجام می‌شود. پارامترهای آماری خطا نیز بعد از اجرای هر شبکه دوباره محاسبه می‌شوند. جدول ۷ تأثیرپذیری هر یک از پارامترهای مؤثر بر ضریب پخش عرضی بر خروجی مدل را نشان می‌دهد.

جدول ۷. آنالیز حساسیت پارامترهای ورودی برای تخمین  $D_T$

مدل	R	RMSE	MAE	OI
کل	۰/۹۶۶	۰/۰۳۷	۰/۰۱۰	۰/۹۵۱
B	۰/۹۲۶	۰/۰۵۴	۰/۰۱۷	۰/۹۰۵
H	۰/۹۶۴	۰/۰۴۴	۰/۰۰۷	۰/۹۳۲
U	۰/۸۷	۰/۰۷۰	۰/۰۱۸	۰/۸۵۱
U*	۰/۹۵	۰/۰۴۳	۰/۰۱۲	۰/۹۳۵

پیشنهادی براساس عملکرد آنها با مجموعه داده‌های تست، ارائه کردند. بر این اساس، حداقل یکی از شیب‌های خط رگرسیون از مبدأ برای مقادیر پیش‌بینی در مقابل مقادیر مشاهده‌ای (K) و یا به عکس (K') باید به عدد یک نزدیک باشد [۳۰].

$$K = \sum_{i=1}^n (O_i \times P_i) / P_i^2 \quad (13)$$

$$K' = \sum_{i=1}^n (O_i \times P_i) / O_i^2 \quad (14)$$

علاوه بر این، ضریب تعیین برای خط رگرسیون از مبدأ باید کمتر از ۰/۱ باشد.

آن از جنس پارامتر خروجی مشاهداتی است. مقدار این شاخص آماری معمولاً مشابه مقدار RMSE است. در جدول ۶ مقادیر تابع هدف به دست آمده برای هر مدل آورده شده است.

جدول ۶. ارزیابی عملکرد مدل‌های پیشنهادی با استفاده از مقادیر تابع هدف

مدل	OBJ
SVM-Poly	۰/۰۲۰
SVM-RBF	۰/۰۱۱
MT	۰/۰۱۷

همان‌طور که مشاهده می‌شود، مدل SVM-RBF مقادیر تابع هدف کمتری نسبت به سایر مدل‌ها در تخمین ضریب پخش عرضی (OBJ=۰/۰۱۱) دارد. از این رو، می‌توان گفت که این مدل دقت بیشتری دارد.

#### آنالیز حساسیت

آنالیز حساسیت (SA)<sup>۱</sup> به مطالعه تأثیرپذیری متغیرهای خروجی از متغیرهای ورودی یک مدل آماری گفته

نتایج جدول ۷ نشان می‌دهد پارامتر سرعت جریان (U) با داشتن بیشترین خطا (RMSE=۰/۰۷۰) و مؤثرترین پارامتر روی ضریب پخش عرضی است. همچنین، عرض کانال با داشتن R=۰/۹۲۶ و شاخص‌های آماری دیگر مربوط به آن (RMSE=۰/۰۵۴، MAE=۰/۰۱۷، OI=۰/۹۰۵)، رتبه دوم را به دست آورد.

#### صحت‌سنجی خارجی برای مدل‌های پیشنهادی

صحت‌سنجی خارجی روشی است به منظور ارزیابی نتایج به دست آمده از روش‌ها که به صورت تصادفی به دست نیامده باشند. تروپشا و همکارانش نوعی معیار صحت‌سنجی خارجی جدیدی برای ارزیابی مدل‌های

$$R_0'^2 = 1 - \sum_{i=1}^n O_i^2 (1 - K')^2 / \sum_{i=1}^n (O_i - \bar{O})^2 \quad (19)$$

معیار صحت‌سنجی و مقادیر هر یک از ضرایب برای مدل‌های پیشنهادی در جدول ۸ نشان داده شده است. همان‌گونه که مشاهده می‌شود، هر سه روش هوشمند براساس معیار صحت‌سنجی خود، عملکردی قابل قبول دارند. بنابراین، این روش‌ها از دقت پیش‌بینی زیادی برخوردار بوده و مقدار همبستگی میان مقادیر پیش‌بینی‌شده و مشاهده‌ای در این روش‌ها، به صورت تصادفی برآورد نشده‌اند.

$$m = (R^2 - R_0^2) / R^2 \quad (15)$$

$$n = (R^2 - R_0'^2) / R^2 \quad (16)$$

که در آن  $R_0^2$  مربع ضریب همبستگی از مبدأ بین مقادیر پیش‌بینی و مشاهده‌ای و  $R_0'^2$  مربع ضریب همبستگی بین مقادیر مشاهده‌ای و پیش‌بینی است که به صورت روابط ۱۷-۱۹ محاسبه می‌شوند [۳۰]:

$$R_m = R^2 \times \left( 1 - \sqrt{|R^2 - R_0^2|} \right) > 0.5 \quad (17)$$

$$R_0^2 = 1 - \sum_{i=1}^n P_i^2 (1 - K)^2 / \sum_{i=1}^n (P_i - \bar{P})^2 \quad (18)$$

جدول ۸. معیار صحت‌سنجی خارجی برای پیش‌بینی ضریب پخش عرضی

مدل	R	K	K'	m	n	R <sub>m</sub>
SVM-Poly	۰/۹۷۲	۰/۶۰۵	۱/۵۶۴	۰/۱۱	۰/۲۹	۰/۶۲۸
SVM-RBF	۰/۹۷۱	۰/۸۹۱	۱/۰۶	-۰/۰۴	-۰/۰۵	۰/۷۴۷
MT	۰/۹۳۴	۰/۹۷۹	۰/۸۹۹	-۰/۱۴۵	-۰/۱۳	۰/۵۶۱

### نتیجه‌گیری

رشد روزافزون جمعیت و صنعتی شدن، سبب افزایش تقاضای آب شده است. با توجه به کمی نزولات جوئی و آب‌های سطحی، شناسایی و کنترل آلودگی رودخانه‌ها که یکی از منابع اصلی تأمین آب هستند، اهمیت زیادی پیدا می‌کند. بنابراین، شناخت و مطالعه فرایندهای اختلاط و انتقال مواد در رودخانه‌ها از جمله فعالیت‌های مهم در برنامه‌های مدیریت منابع آب به شمار می‌آید. بین فرایندهای اختلاط، پس از پدیده انتشار طولی، فرایند انتشار عرضی آلودگی، تأثیرگذارترین پارامتر محسوب می‌شود. از این رو، هدف از پژوهش حاضر پیش‌بینی ضریب پخش عرضی انتقال آلودگی در جریان‌های سطحی است که برای این منظور ابتدا با استفاده از مدل درخت (MT) و ماشین بردار پشتیبان (SVM) به پیش‌بینی ضریب پخش عرضی انتقال آلودگی در رودخانه پرداخته شده است. به این منظور، از ۱۸۷ سری داده که شامل عمق جریان، سرعت جریان، سرعت برشی و عرض کانال که از مطالعات گذشته جمع‌آوری شده، استفاده شده است. نتایج به‌دست آمده از معیارهای ارزیابی نشان داد مدل

SVM-RBF توانایی بیشتری در تخمین ضریب پخش عرضی داشت. علاوه بر این، مقایسه عملکرد روش‌های هوشمند با روابط تجربی بیان می‌کند که روابط تجربی نتوانستند دقت قابل قبولی را از خود نشان دهند. همچنین، نتایج بررسی تأثیر هر یک از پارامترهای ورودی بر مقدار  $D_T$  از آنالیز حساسیت که توسط مدل SVM-RBF انجام شد، نشان داد که سرعت جریان بیشترین تأثیر را بر ضریب پخش عرضی دارد.

### منابع

- [1]. Riahi-Madvar H, Ayyoubzadeh SA, Khadangi, E, Ebadzadeh, MM. An expert system for predicting longitudinal dispersion coefficient in natural streams by using ANFIS. Expert Systems with Applications. 2009; 36(4), pp.8589-8596.
- [2]. Fischer, HB, List E, Koh R, Imberger J, Brooks N. Mixing in inland and coastal waters Academic Press. New York, 1979.
- [3]. Chau KW. Transverse mixing coefficient measurements in an open rectangular channel. Advances in Environmental Research. 2000; 4(4), 287-294.

- [4]. Fischer HB. Transverse mixing in a sand-bed channel. Professional Paper. 1967; No. 575-D, U.S. Geological Survey
- [5]. Rutherford JC. River Mixing. John Wiley and Sons, Chichester, UK. 1967.
- [6]. Deng ZQ, Singh VP, Bengtsson L. Longitudinal dispersion coefficient in straight rivers. *J Hydraul Eng*. 2001; 127:919–927
- [7]. Bansal MK (1971) Dispersion in natural streams. *J Hydr Div ASCE* 97(11): 1867–1886.
- [8]. Kisi O, Parmar KS. Application of least square support vector machine and multivariate adaptive regression spline models in long term prediction of river water pollution, *Journal of Hydrology*. 2015; 534, 104-112.
- [9]. Khosravi A, Koury R, Machado L, Pabon, J. Prediction of wind speed and wind direction using artificial neural network, support vector regression and adaptive neuro-fuzzy inference system. *Sustainable Energy Technologies and Assessments*. 2018; 25, 146-160.
- [10]. Najafzadeh M, Ghaemi A, Emamgholizadeh S. Prediction of water quality parameters using evolutionary computing-based formulations. *International Journal of Environmental Science and Technology*. 2018; 1-20.
- [11]. Sahay RR. Prediction of longitudinal dispersion coefficients in natural rivers using artificial neural network, *Environ, Fluid Mech*. 2011; 11(3), 247-261.
- [12]. Etemad-Shahidi A, Taghipour M. Predicting longitudinal dispersion coefficient in natural streams using M5' model tree. *Journal of Hydraulic Engineering, ASCE*. 2012; 138(6), 542-554.
- [13]. Najafzadeh M, Tafarjoruz A. Evaluation of neuro-fuzzy GMDH-based particle swarm optimization to predict longitudinal dispersion coefficient in rivers. *Environmental Earth Sciences*. 2016; 75(2), 157.
- [14]. Rezaei-Balf M, Noori R, Berndtsson R, Ghaemi A, Ghiasi B. Evolutionary polynomial regression approach to predict longitudinal dispersion coefficient in rivers. *Journal of Water Supply: Research and Technology-Aqua*. 2018; 67(5), 447-457.
- [15]. Huai W, Shi H, Yang Z, Zeng Y. Estimating the Transverse Mixing Coefficient in Laboratory Flumes and Natural Rivers. *Water, Air, & Soil Pollution*. 2018; 229(8): p. 252.
- [16]. Vapnik VN, *Statistical Learning Theory*. John Wiley. New York, 1998; 16-29.
- [17]. Noori R, Abdoli MA, Ameri A, Jalili-Ghazizade M. Prediction of municipal solid waste generation with combination of support vector machine and principal component analysis: A case study of Mashhad. *Environmental Progress and Sustainable Energy*. 2008; 28 (2), 249-258.
- [18]. Quinlan JR. Learning with continuous classes. In: *Proceedings of the Fifth Australian Joint Conference on Artificial Intelligence*, World Scientific. 1992; 343-348.
- [19]. Zahiri A, Azamathulla HMd. Comparison between linear genetic programming and M5 tree models to predict flow discharge in compound channels. *Neural Computing and Applications*. 2012; 24(2), 413-420.
- [20]. Rahimikhoob A. Comparison of M5 Model Tree and Artificial Neural Network's Methodologies in Modelling Daily Reference Evapotranspiration from NOAA Satellite Images. *Water Resources Management*. 2016; 1-13.
- [20]. Jeon TM, Baek KO, Seo IW. Development of an empirical equation for the transverse dispersion coefficient in natural streams. *Environmental Fluid Mechanics*. 2007; 7(4), 317– 329.
- [21]. Tabatabaei SH, Heidarpour M, Ghasemi M, Hoseinipour EZ. Transverse mixing coefficient on dunes with vegetation on a channel wall. *World Environmental and Water Resources Congress*. 2013; 1903–1911.
- [22]. Long T, Guo J, Feng Y, Huo G. Modulus of transverse diffuse simulation based on artificial neural network. *Chongqing Environmental Science*. 2002; 24(2), 25–28.
- [23]. Pilechi A, Mohammadian A, Rennie CD, Zhu DZ. Efficient method for coupling field data and numerical modeling for the estimation of transverse mixing coefficients in meandering rivers. *Journal of Hydraulic Engineering*. 2016; 142(6), 04016009.
- [24]. Beltaos S, Day TJ. A field study of longitudinal dispersion. *Canadian Journal of Civil Engineering*. 1974; 5(4), 572–585.
- [25]. Webel, G, Schatzmann M. Transverse mixing in open channel flow. *Journal of Hydraulic Engineering*. 1978; 110(4), 423–435.
- [26]. Yotsukura, N, Sayre WW. Transverse mixing in natural channels. *Water Resources Research*. 1976; 12(4), 695–704.

- [27]. Engmann JEO, Kellerhals R. Transverse mixing in an ice-covered river. *Water Resources Research*. 1974; 10(4), 775–784.
- [28]. Najafzadeh M, Ghaemi A. Prediction of the five-day biochemical oxygen demand and chemical oxygen demand in natural streams using machine learning methods. *Environmental monitoring and assessment*. 2019; 191(6), 380.
- [29]. Gandomi AH, Alavi AH, Sahab MG, Arjmandi P. Formulation of elastic modulus of concrete using linear genetic programming. *Journal of Mechanical Science and Technology*. 2010; 24(6), 1273-1278.
- [30]. Gandomi A, Yun G, Alavi A. An evolutionary approach for modeling of shear strength of RC deep beams. *Materials and Structures*. 2013; 46,2109-2119.