

هیدروژئومورفولوژی، شماره ۱۰، بهار ۱۳۹۶، صص ۱۴۳-۱۲۱

وصول مقاله: ۱۳۹۵/۰۷/۱۷ تأیید نهایی مقاله: ۱۳۹۶/۰۲/۱۸

بررسی اثر مقیاس زمانی (روزانه، ماهانه و سالانه) در پیش‌بینی بار رسوب

معلق

مریم اسدی^۱

علی فتح‌زاده^{۲*}

روح‌الله تقی‌زاده مهرجردی^۳

چکیده

تعیین بار رسوبی معلق رودخانه‌ها یکی از پروژه‌های مهم مهندسی رودخانه می‌باشد. پیش‌بینی بار رسوبی معلق کمک شایانی در زمینه مدیریت منابع آبی خواهد نمود. سوال اصلی در این تحقیق بررسی نقش انواع داده‌های روزانه، ماهانه و سالانه در برآورد بار رسوبی با استفاده از مدل‌های مبتنی بر یادگیری ماشین است. بدین منظور از آمار بار معلق در سه مقیاس زمانی روزانه، ماهانه و سالانه ایستگاه هیدرومتری اوهایو واقع در ایالات متحده آمریکا در فاصله سال‌های ۱۹۹۲ تا ۲۰۱۴ استفاده گردید. با هدف انتخاب مناسب‌ترین مدل، مدل‌های شبکه عصبی مصنوعی پس انتشار خطا و تابع پایه‌ی شعاعی، رگرسیون خطی، k نزدیک‌ترین همسایه، درخت تصمیم $M5$ ، فرآیند گوسی، ماشین بردار پشتیبان و ماشین بردار پشتیبان تکاملی اجرا و مورد ارزیابی قرار گرفتند. نتایج حاصل از این تحقیق نشان داد که برای داده‌های روزانه، مدل k نزدیک‌ترین همسایه با $RMSE=5/28$ ؛ برای داده‌های ماهانه مدل فرآیند گوسی با $RMSE=8/7$ و برای داده‌های

۱- دانشجوی کارشناسی ارشد آبخیزداری، دانشکده کشاورزی و منابع طبیعی دانشگاه اردکان، اردکان

۲- دانشیار دانشکده کشاورزی و منابع طبیعی دانشگاه اردکان، اردکان (نویسنده مسئول)

Email: fatzade@ardakan.ac.ir

۳- استادیار دانشکده کشاورزی و منابع طبیعی دانشگاه اردکان، اردکان

سالانه مدل فرآیند گوسی با مقدار $RMSE=7/2$ مناسب‌ترین مدل جهت پیش‌بینی بار رسوبی معلق بوده‌اند. همچنین مقایسه مقادیر ارزیابی مدل‌ها حاکی از آن است که پیش‌بینی داده‌های سالانه نسبت به ماهانه و روزانه از دقت بیشتری برخوردار هستند.

کلمات کلیدی: بار معلق، شبکه عصبی مصنوعی، فرآیند گوسی، ماشین بردار پشتیبان، ماشین بردار تکاملی.

مقدمه

پدیده رسوب یکی از عوامل ایجاد بحران‌های کمی و کیفی آب‌های سطحی محسوب می‌شود. رسوبات حمل شده توسط جریان، حجم آب خالص ورودی به مخازن را تحت تأثیر قرار داده و با گذشت زمان از حجم مخازن و سدها می‌کاهد. برآورد بار رسوبی معلق که جریان مشخص قادر به حمل آن باشد، از موضوعات اصلی تحقیقات رسوب بوده که در بسیاری از پروژه‌های مهندسی همچون برنامه‌ریزی و طراحی منابع ذخیره‌ی آب، برآورد بار معلق سالانه برای آبیگرهای رودخانه، طراحی و نگهداری کانال‌های آبیاری پایدار، حفاظت سواحل و لایروبی کانال‌ها حائز اهمیت است (کیگیزگلو^۱، ۲۰۰۳: ۱۸۵). بار رسوب معلق با اندازه‌گیری مستقیم و معادله انتقال رسوب تعیین می‌گردد.

اگرچه اندازه‌گیری بار رسوبی رودخانه قابل اعتمادتر است اما این کار هزینه‌بر بوده و برای تمام طول رودخانه میسر نمی‌باشد. معادلات انتقال رسوب نیز نیاز به اطلاعات دقیق‌تری از جریان و بستر رودخانه داشته که گردآوری آنها نیاز به انجام عملیات صحرایی فراوانی دارد (ازتورک و همکاران^۲، ۲۰۰۱: ۶۴۴). لذا به کارگیری

1- Cigizoglu

2- Ozturk et al.,

روش‌هایی که به طور غیرمستقیم و با استفاده از پارامترهای مؤثر در بار رسوب، بتوان آن را برآورد نمود کمک شایانی در زمینه‌ی پیش‌بینی بار رسوب خواهد کرد.

روش‌های داده‌ی کاوی ابزارهای هوشمندی برای برقراری این رابطه‌ی غیرخطی می‌باشند (آلپ و کیگیزگلو^۱، ۲۰۰۷: ۲). روش‌های داده‌کاوی یکی از روش‌هایی است که امروزه در مطالعات فرآیندهای هیدرولوژی و اقلیمی بسیار مورد استفاده قرار گرفته‌است و ابزار هوشمند به منظور برقراری ارتباط بین پارامترهای اقلیمی و هیدرولوژیکی است (چنگ و همکاران^۲، ۲۰۱۵: ۸۷؛ رسولی و کانون^۳، ۲۰۱۲: ۲۸۴؛ بهتچریا و سلمتین^۴، ۲۰۰۵: ۳۸۱) با وجود توانمندی بالای تکنیک‌های یادگیری ماشین به نظر می‌رسد پایه‌ی زمانی انواع داده‌های رسوب رفتار متفاوتی را نسبت به مدل‌های شبیه‌سازی خود نشان می‌دهند. تا کنون در مطالعات مختلف به بررسی و مدل‌سازی بار رسوبی در سری‌های زمانی روزانه، ماهانه و سالانه پرداخته شده است، به طور مثال کومار و همکاران^۵ (۲۰۱۶) از بین سری‌های زمانی، داده‌های روزانه را انتخاب و به برآورد بار رسوب معلق روزانه با استفاده از روش‌های یادگیری ماشین پرداختند. برای این منظور، از مدل‌های شبکه‌ی عصبی مصنوعی، شبکه‌های عصبی تابع پایه‌ی شعاعی، حداقل ماشین‌بردار پشتیبانی مربع، رگرسیون چند خطی و مدل درخت تصمیم، مانند طبقه‌بندی و درخت رگرسیون و مدل درختی M5 استفاده گردید. نتایج نشان داد که تمام مدل‌ها قادر به شبیه‌سازی رسوب معلق می‌باشند و مدل حداقل ماشین‌بردار پشتیبانی مربع با دقت ۰/۸۹ به خوبی قادر به مدل‌سازی رسوب معلق روزانه می‌باشد.

1- Alp & Cigizoglu

2- Chang et al.,

3- Rasouli & Cannon

4- Bhattacharya & and Solomatine

5- Kumar et al.,

ذونعمت کرمانی و همکاران^۱ (۲۰۱۶) به مدل سازی غلظت رسوب معلق روزانه براساس مدل های داده محور پرداختند. آنان نیز در مطالعه ی خود مدل های شبکه ی عصبی مصنوعی، ماشین بردار رگرسیون و منحنی سنجی رسوب مورد مقایسه قرار دادند. نتایج براساس معیارهای ارزیابی در این مدل ها بیانگر آن بود که مدل ماشین بردار رگرسیون با تابع پایه ی شعاعی دارای عملکرد مناسب تری به منظور پیش بینی بار رسوبی معلق می باشد.

در مطالعه ای دیگر لافدانی و همکاران^۲ (۲۰۱۳) از شبکه های عصبی مصنوعی و ماشین بردار پشتیبان به منظور پیش بینی بار رسوب معلق روزانه استفاده نمودند. نتایج به دست آمده نشان داد که مدل ماشین بردار رگرسیون با استفاده از آزمون گاما دارای عملکرد مناسبی در پیش بینی بار رسوبی معلق می باشد. مدل سازی رسوبات معلق ماهانه نیز مطالعه ی یوسفی و پورشرعیاتی (۱۳۹۲) مورد بررسی قرار گرفت. آنان با بررسی سه شبکه Fitting و Forward Back prop Cascade و پیشخور پس انتشار خطا به تعیین الگوریتم مناسب جهت برآورد بار معلق پرداختند. آنان در نتایج خود بیان نموده اند که از میان سه شبکه، شبکه ی عصبی Fitting نتایج بهتری را ارائه داده است.

مقایسه سری های زمانی و فصلی در مطالعه ی کیا و عمادی (۱۳۹۲) مورد توجه قرار گرفت. نتایج مدل سالانه نشان داد که این مدل نسبت به سایر حالات به مدل مناسب این ایستگاه نزدیک تر است، البته در این مطالعه تفکیک زمانی فصلی داده ها به دو فصل سیلابی و غیرسیلابی نشان داد که با تفکیک داده ها مدل ها نتایج بهتری را ارائه می نمایند.

1- Zounemat-Kermani et al.,

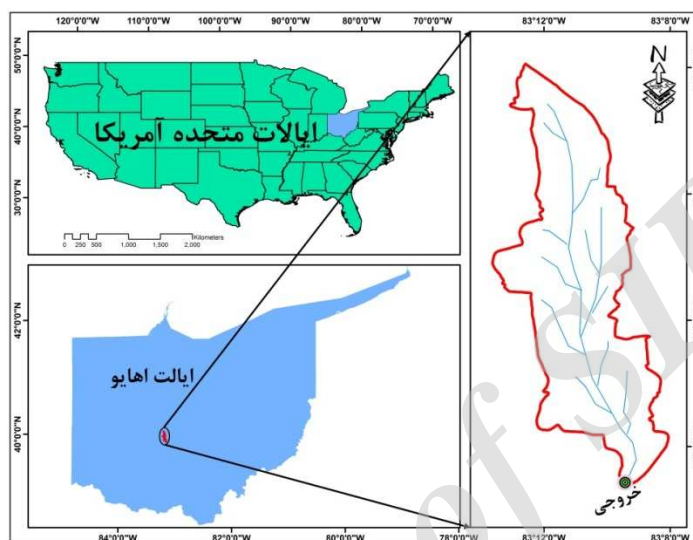
2- Lafdani et al.,

آنچه مشخص است این است که دقت مدل‌ها در پیش‌بینی بار رسوبی در سری‌های زمانی با توجه به حجم داده‌ها متفاوت بوده و مدل بر روی هر یک از سری‌های زمانی نتایج متفاوتی را ارائه می‌دهد. در این مطالعه با تهیه‌ی داده‌های دبی و رسوب معلق مربوط به سری‌های زمانی روزانه، ماهانه و سالانه در یک حوضه‌ی ثابت به بررسی میزان دقت مدل‌ها در پیش‌بینی بار رسوبی معلق در سری‌های زمانی روزانه، ماهانه و سالانه پرداخته شده‌است. از طرفی با توجه به گستردگی مدل‌های یادگیری ماشین در این مطالعه به بررسی توانمندی هشت مدل متفاوت پرداخته شد و قابلیت آن‌ها با استفاده از معیارهای مناسب مورد ارزیابی قرار گرفت.

مواد و روش‌ها

معرفی منطقه‌ی مورد مطالعه

منطقه‌ی مطالعاتی ایستگاه هیدرومتری واقع در ایالت اوهایو می‌باشد. مساحت حوضه ۷۲/۹۲ کیلومتر مربع بوده و دوره‌ی آماری مورد نظر از سال ۱۹۹۲ تا سال ۲۰۱۳ (۲۱ سال) بوده که تعداد ۷۸۷۳ داده‌ی رسوب و دبی در این مدت برداشت گردیده است (USGS). متوسط دبی عبوری از این رودخانه ۱/۰۹ مترمکعب بر ثانیه و متوسط رسوب حمل شده ۹/۲۳ میلی‌گرم بر لیتر می‌باشد (شکل ۱).



شکل (۱) حوضه و ایستگاه مورد مطالعه واقع در ایالت اوهایو

در این مطالعه از داده‌های دبی و رسوب روزانه، ماهانه و سالانه برداشت شده در ایستگاه هیدرومتری در سرشاخه‌های رودخانه‌ی اوهایو استفاده شده است.

مدل‌های مورد استفاده

در این مرحله داده‌های آماده شده با استفاده از تکنیک‌ها و عملیات داده‌کاوی مورد کاوش قرار گرفتند تا الگوهای مورد نظر کشف گردند. این مدل‌ها عبارتند از: مدل‌های رگرسیون خطی، شبکه‌های عصبی پیش‌خور پس انتشار و تابع پایه‌ی شعاعی، K نزدیک‌ترین همسایه، ماشین بردار پشتیبان، ماشین بردار پشتیبان تکاملی و مدل درخت تصمیم M5.

– شبکه‌ی عصبی پس انتشار خطا

شبکه‌ی عصبی پس انتشار خطا حداقل شامل سه لایه‌ی ورودی، خروجی و لایه‌ی

مخفی است. هر نرون در یک لایه وزن ورودی را از لایه قبلی دریافت کرده آن را به عنوان خروجی نرون در لایه بعدی انتقال می‌دهد. هورنیک^۱، ۱۹۸۹: ۳۵۹. نتایج شبکه با نتایج واقعی مقایسه شده و خطای شبکه توسط معادله محاسبه می‌شود. روند آموزش تا زمانی که خطا به مقدار قابل قبولی برسد ادامه می‌یابد.

$$Y_{net} = \sum_{i=1}^n X_i \cdot W_i + W_o \quad \text{رابطه (۱)}$$

$$Y = f(Y_{net}) = \frac{1}{1 + e^{-Y_{net}}} \quad \text{رابطه (۲)}$$

$$J_r = \frac{1}{2} \cdot \sum_{i=1}^k (Y_i - O_i)^2 \quad \text{رابطه (۳)}$$

در این روابط Y_i : پاسخ نرون i ، $f(Y_{net})$: تابع متحرک غیرخطی، Y_{net} : مجموع وزن‌های ورودی، X_i : نرون ورودی، W_i : وزن هر نرون ورودی، O_i : مقدار مشاهده شده نرون i ، W_o : بایاس، J_r : خطای بین مقدار مشاهداتی و پاسخ شبکه می‌باشد (روشگر و پرهیز جوان، ۱۳۹۳: ۱۷۸).

– شبکه‌ی عصبی با تابع پایه‌ی شعاعی

شبکه‌ی عصبی تابع پایه‌ی شعاعی به عنوان یک شبکه‌ی سه‌لایه که در آن لایه‌ی پنهان به صورت ثابت واجد تابع تبدیل غیرخطی با پارامترهای غیرقابل تنظیم است، شناخته می‌شود. برای هر گره، فاصله‌ی اقلیدسی میان مرکز و بردار ورودی به وسیله‌ی تابع غیرخطی که خروجی گره‌ها در لایه‌ی پنهان را تخمین می‌زند، محاسبه می‌شود. سپس لایه‌ی خروجی این نتایج را در یک نگاشت خطی ترکیب می‌کند.

1- Hornik

K- نزدیک‌ترین همسایه^۱

الگوریتم k نزدیک‌ترین همسایه شامل انتخاب تعداد مشخصی از بردار داده‌ها و سپس باز نمونه‌گیری تصادفی از آن مجموعه برای شبیه‌سازی دوره‌ی زمانی متعاقب یک دوره‌ی معین می‌باشد. از مزایای استفاده از این الگوریتم در پیش‌بینی‌ها می‌توان به اجرای ساده، عدم نیاز به مرحله‌ی تخمین پارامترها، قابلیت مدل‌سازی غیرخطی، مؤثر بودن و عملکرد با بازدهی بالا در برخورد با تعداد دسته‌های زیاد از داده‌ها می‌باشد (عزمی و عراقی‌نژاد، ۱۳۹۱: ۱۱۰).

- مدل درختی M5p

این مدل زیر مجموعه‌ای از روش‌های یادگیری ماشینی و داده‌کاوی است. ساختار یک مدل درختی شامل ریشه، گره‌های داخلی و برگ می‌باشد. برای اولین بار کوینلان (۱۹۹۲) مدل درخت تصمیم موسوم به M5 را برای پیش‌بینی داده‌های پیوسته ارائه نمود. این مدل، بر خلاف مدل‌های درخت تصمیم معمول که کلاس یا رده‌های گسسته را به عنوان خروجی ارائه می‌کنند، یک مدل خطی چندمتغیره را برای داده‌ها در هر گره از مدل درختی می‌سازد. تشکیل ساختار مدل‌های درخت تصمیم‌گیری شامل مراحل ایجاد درخت و هرس کردن آن است (ویتن و فرانک، ۲۰۰۵). معیار تقسیم برای الگوریتم مدل M5 ارزیابی انحراف معیار مقادیر کلاسی است که به عنوان کمیتی از خطا به یک گره می‌رسد و کاهش مورد انتظار در این خطا را به عنوان نتیجه آزمون هر صفت در آن گره محاسبه می‌نماید. کاهش انحراف معیار (SDR) از رابطه‌ی (۴) به دست می‌آید:

$$SDR = sd(T) - \sum \frac{|T_i|}{|T|} sd(T_i) \quad (۴) \text{ رابطه‌ی}$$

1- K Nearest Neighbor

که در آن T ، بیانگر یک سری نمونه‌هایی است که به گره می‌رسد. T_i بیانگر نمونه‌هایی است که i امین خروجی سری پتانسیلی را دارند و sd بیانگر انحراف معیار است (آلبرگ و همکاران، ۲۰۱۲: ۷۱). به دلیل فرآیند انشعاب، داده‌های قرار گرفته در گره‌های فرزند، انحراف معیار کم‌تری نسبت به گره مادر داشته و بنابراین خالص‌تر هستند. پس از حداکثرسازی تمامی انشعاب‌های ممکن $M5$ صفتی را انتخاب می‌کند که کاهش مورد انتظار را بیشینه نماید (ظهیری و قربانی، ۱۳۹۲: ۱۲۱).

– فرآیند گوسی^۱ (GP)

فرآیند گوسی یک روش یادگیری ماشینی غیرپارامتری قوی برای ایجاد مدل‌های احتمال‌گرایانه جامع از مسائل دنیای واقعی است. به طور رسمی، یک فرآیند گوسی تولید داده‌هایی می‌کند که در طول این طیف قرار دارد به طوری که هر زیرمجموعه متناهی دامنه‌ی یک توزیع گوسی چندمتغیره را دنبال می‌کند.

– ماشین بردار پشتیبان^۲ (SVM)

روش ماشین بردار پشتیبان از روش‌های یادگیری ماشینی است که بر مبنای تئوری یادگیری آماری $vapnik$ در دهه‌ی ۹۰ میلادی توسط $Vapnik$ و همکاران ارائه شد. در ماشین بردار پشتیبان از اصول کمینه کردن ریسک ساختاری استفاده شده؛ در حالی که سایر روش‌ها از اصول کمینه کردن تجربی بهره می‌برند (نظری و همکاران^۳، ۲۰۱۱).

1- Gaussian Process

2- Support Vector Machine

3- Nazari et al.,

این روش بر پایه‌ی نظریه‌ی یادگیری محاسباتی توسعه‌یافته است (هانگ و چنگ، ۲۰۰۷: ۵۱۶). در حالتی که داده‌ها به صورت خطی به شکل رابطه‌ی (۵) قابل ارائه‌اند:

$$y = \text{sign} \left[\sum_{i=1}^N y_i \alpha_i (X * X_i) + b \right] \quad \text{رابطه‌ی (۵)}$$

در رابطه‌ی (۵)، پارامترهای ابر صفحه هستند. اگر داده‌ها به صورت خطی تفکیک‌پذیر نباشند، رابطه‌ی (۵) به رابطه‌ی (۶) تغییر می‌یابد:

$$y = \text{sign} \left[\sum_{i=1}^N y_i \alpha_i K(X, X_i) + b \right] \quad \text{رابطه‌ی (۶)}$$

$K(X, X_i)$ تابع کرنلی است که، برای ایجاد تابع ماشین‌هایی با انواعی مختلف از سطوح تصمیم‌گیری غیرخطی در فضای داده‌ها، ضرب‌های داخلی تولید می‌کند.

– ماشین بردار پشتیبان تکاملی (ESVM)

مدل ماشین بردار تکاملی یک استراتژی تکاملی را به‌منظور بهینه‌سازی به کار می‌برد. در واقع این مدل یک الگوریتم تکاملی را برای حل مشکل بهینه‌سازی دوگانه ارائه می‌دهد. ESVM با تجزیه و تحلیل فرکانس از ویژگی‌های انتخاب شده، یک مجموعه‌ی کوچک از ویژگی‌های مؤثر را شناسایی کرده و از این رو باعث بهبود دقت طبقه‌بندی می‌گردد (هانگ و چنگ، ۲۰۰۷: ۵۱۷).

در مدل‌های فوق، فرآیند مدل‌سازی بار رسوبی معلق در سری‌های زمانی روزانه، ماهانه و سالانه انجام شد. بدین ترتیب که مجموعه‌ی داده‌های دبی روزانه، ماهانه و

سالانه به عنوان ورودی مدل به نرم‌افزار ریپید ماینر وارد شد و سپس مقادیر بار رسوبی معلق در هر یک از بازه‌های زمانی برآورد گردید.

– ارزیابی مدل

یک روش جایگزین برای زیرنمونه‌برداری تصادفی روش Cross-Validation است. در این روش هر رکورد داده‌ای یک بار در مرحله آموزش و یک بار در مرحله تست استفاده می‌گردد. حالت عمومی‌تر این روش k-fold cross-validation نام دارد که آن مقدار k برابر N، یعنی برابر اندازه مجموعه داده‌های اولیه قرار می‌گیرد. این حالت Leave-One-Out نامیده می‌شود. مزیت این حالت در آن است که از بیشترین تعداد داده‌های ممکن برای مرحله آموزش استفاده می‌کند. (صنعی آباد و همکاران، ۱۳۹۳: ۱۹۲). در این مطالعه نیز به منظور ارزیابی مدل از روش Leave-One-Out استفاده شد.

به منظور ارزیابی نتایج از معیارهای ارزیابی Root mean squared error (RMSE) و Correlation coefficient (r) استفاده شد.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (o_i - p)^2}{n}} \quad \text{رابطه‌ی (۷)}$$

$$r = \frac{\sum_{i=1}^n (p_i - \bar{p})(o_i - \bar{o})}{\sqrt{\sum_{i=1}^n (p_i - \bar{p})^2 \sum_{i=1}^n (o_i - \bar{o})^2}} \quad \text{رابطه‌ی (۸)}$$

که در این معادلات o_i مقادیر مشاهده‌ای، p_i داده‌های پیش‌بینی شده، \bar{p} میانگین داده‌های پیش‌بینی شده، \bar{o} میانگین داده‌های مشاهده شده، n تعداد داده‌ها می‌باشد. با توجه به مقادیر بار رسوبی واقعی و مقادیر پیش‌بینی شده توسط مدل‌های اعمال شده فرآیند ارزیابی مدل‌ها صورت گرفت.

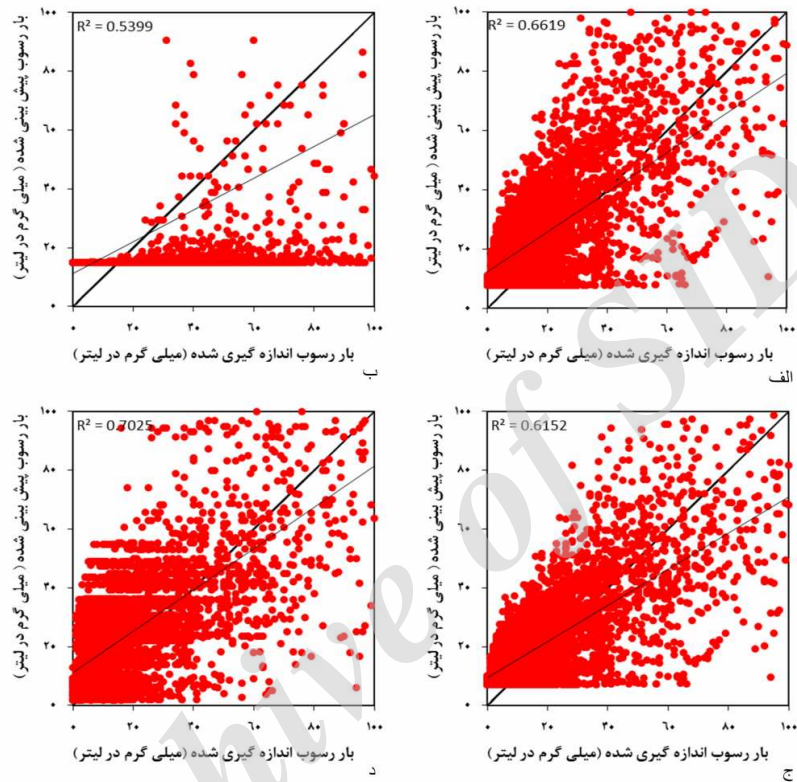
بحث و نتایج

خلاصه‌ی آماری داده‌های دبی و رسوب در مقیاس زمانی مختلف در جدول (۱) ارائه شده است.

جدول (۱) خلاصه‌ی آماری داده‌های دبی (فوت مکعب بر ثانیه) و رسوب (میلی‌گرم بر لیتر)

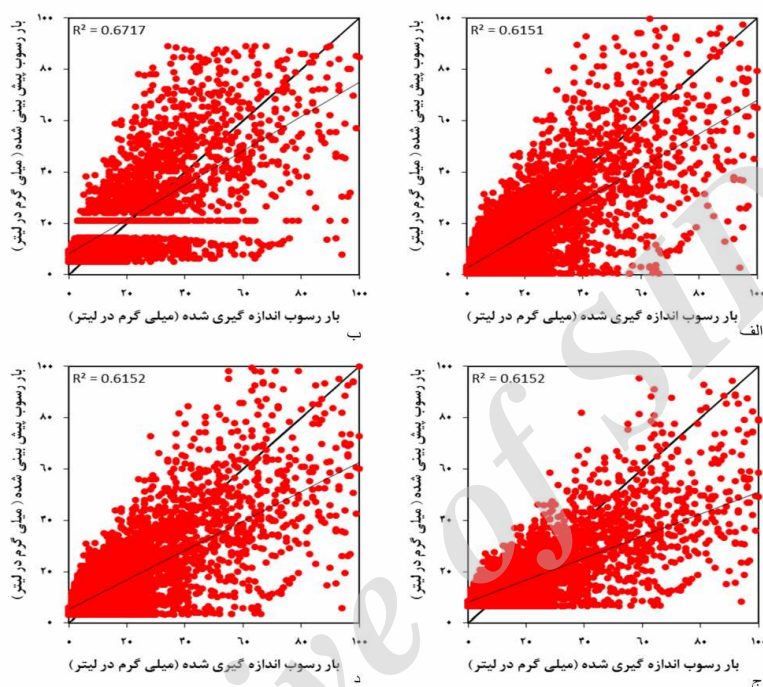
سری زمانی	نوع داده	حداقل	حداکثر	میانگین	انحراف از معیار	ضریب تغییرات
روزانه	دبی	۰/۰۱	۲۰۰۰	۴۱/۶	۹۴/۶	۲/۳
	رسوب	۰/۴	۸۱۹	۲۵/۴	۵۲/۳	۲/۱
ماهانه	دبی	۰/۰۰۵	۴/۳۰۷	۳۹/۷	۴۳/۴	۱/۰
	رسوب	۷/۰	۱۰۰	۲۴/۴	۵/۱۹	۰/۸۰
سالانه	دبی	۵/۱۸	۷۰/۹	۳۹/۰	۱۳/۵	۰/۳
	رسوب	۹/۱۲	۴۲/۸	۲۴/۳	۷/۱	۰/۳

نتایج کلی حاکی از توانمندی تمامی مدل‌ها در پیش‌بینی بار رسوبی معلق است. این در جالی است که مدل K نزدیک‌ترین همسایه با مقدار $RMSE=28/5$ و $r=0/83$ و پس از آن مدل درختی M5 بهترین مدل پیش‌بینی بار رسوبی معلق روزانه می‌باشد (شکل‌های ۲ و ۳).



MODEL	FFNN	RBFNN	KNN	LR
RMSE	۳۹/۳	۳۵/۹	۲۸/۵	۳۳
R	۰/۸	۰/۷	۰/۸۳	۰/۸

شکل (۲) نمودار پراکندگی مقادیر مشاهده شده و پیش‌بینی شده بار رسوب معلق روزانه: توسط مدل‌های الف: شبکه عصبی مصنوعی پس انتشار خطا، ب: شبکه‌ی عصبی تابع پایه‌ی شعاعی، ج: رگرسیون خطی، د: k نزدیک‌ترین همسایه

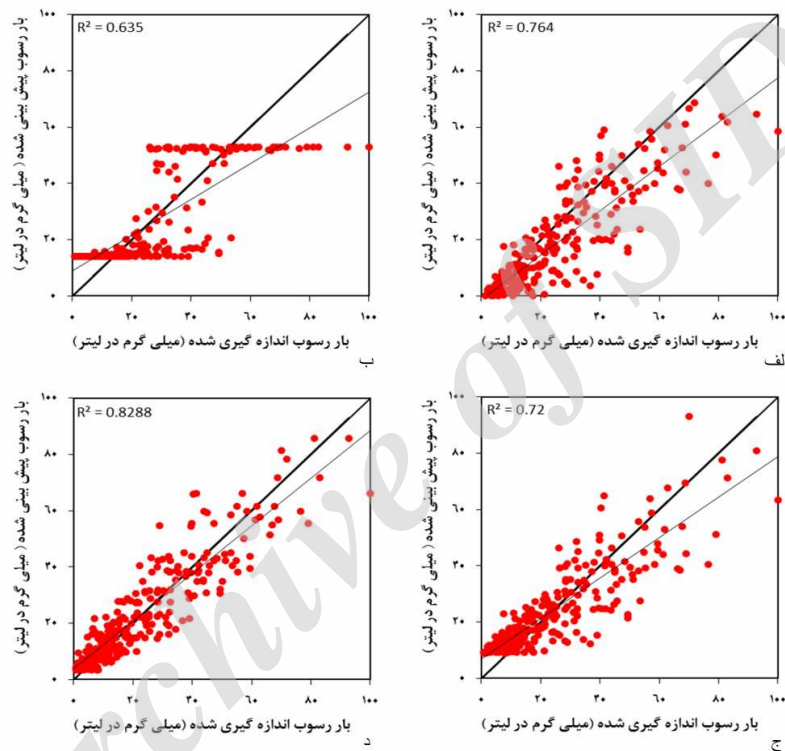


MODEL	GP	M5	ESVM	SVM
RMSE	۳۱/۹	۲۹/۷	۳۲/۶	۳۴/۹
r	۰/۸	۰/۸	۰/۷۸	۰/۸

شکل (۳) نمودار پراکندگی مقادیر مشاهده شده و پیش‌بینی شده بار رسوب معلق روزانه: توسط مدل‌های الف: فرآیند گوسی، ب: درخت تصمیم M5، ج: ماشین بردار پشتیبان، د: ماشین بردار پشتیبان تکاملی

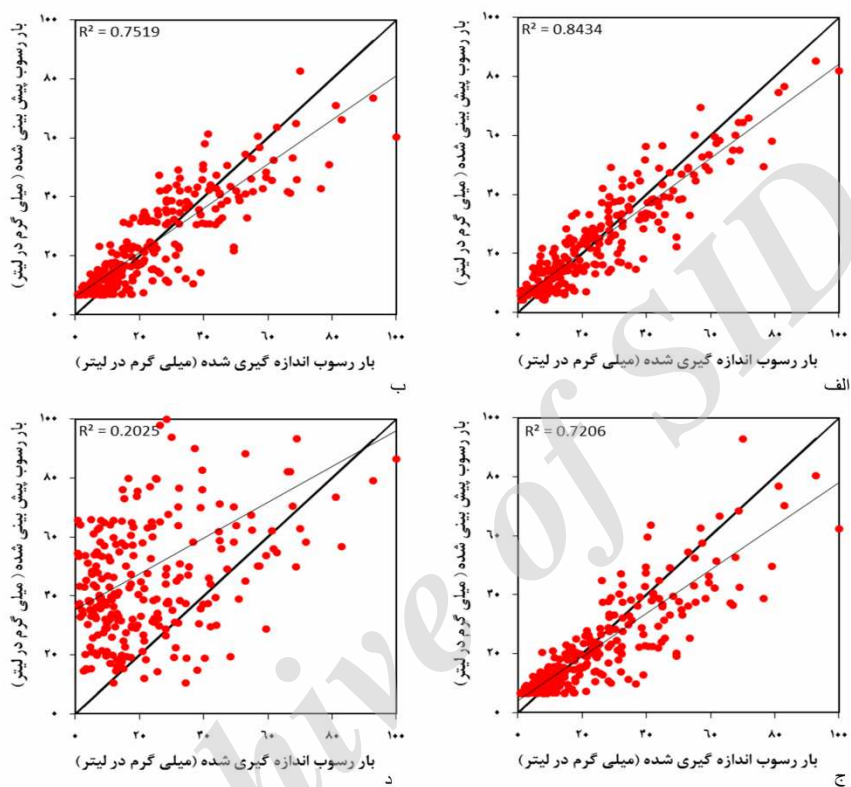
نتایج اعمال مدل‌ها بر روی سری زمانی ماهانه نیز نشان داد که مدل فرآیند گوسی با مقدار $RMSE = 7/8$ و $R = 0/91$ و K نزدیک‌ترین همسایه با مقدار با مقدار $RMSE = 1$ و $R = 0/91$ دارای بیشترین دقت در پیش‌بینی بار رسوب معلق ماهانه می‌باشد. همچنین نتایج نشان می‌دهد که دقت و همبستگی تمامی مدل‌ها نیز نسبت به سری روزانه افزایش پیدا کرده است. مقادیر پراکنش داده‌های واقعی و

پیش‌بینی شده داده‌های ماهانه و نتایج ارزیابی انواع مدل‌ها در شکل‌های ۴ و ۵ نشان داده شده است.



MODEL	FFNN	RBFNN	KNN	LR
RMSE	۱۱/۴	۱۱/۸	۸/۱	۳/۱۰
r	۰/۸۷	۰/۷۹	۰/۹۱	۰/۸۴

شکل (۴) نمودار پراکندگی مقادیر مشاهده شده و پیش‌بینی شده بار رسوب معلق ماهانه: توسط مدل‌های الف: شبکه‌ی عصبی مصنوعی پس انتشار خطا، ب: شبکه‌ی عصبی تابع پایه‌ی شعاعی، ج: رگرسیون خطی، د: k نزدیک‌ترین همسایه

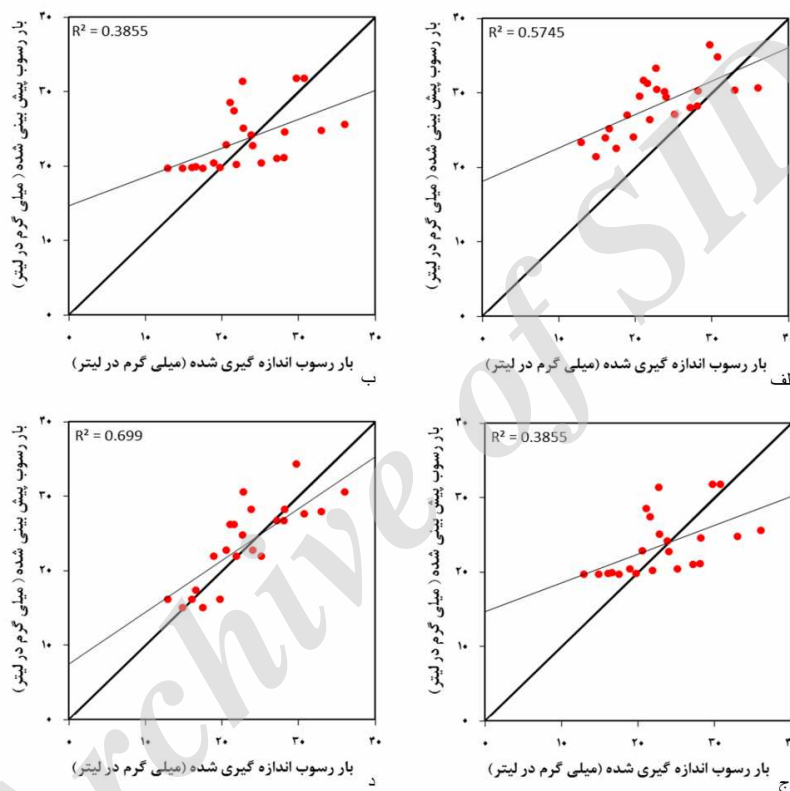


MODEL	GP	M5	SVM	ESVM
RMSE	۸/۷	۹/۷	۱۰/۶	۷/۳۵
r	۰/۹۱	۰/۸۶	۰/۸۴	۰/۴۵

شکل (۵) نمودار پراکندگی مقادیر مشاهده شده و پیش‌بینی شده بار رسوب معلق ماهانه: توسط مدل‌های الف: فرآیند گوسی، ب: درخت تصمیم M5، ج: ماشین بردار پشتیبان، د: ماشین بردار پشتیبان تکاملی

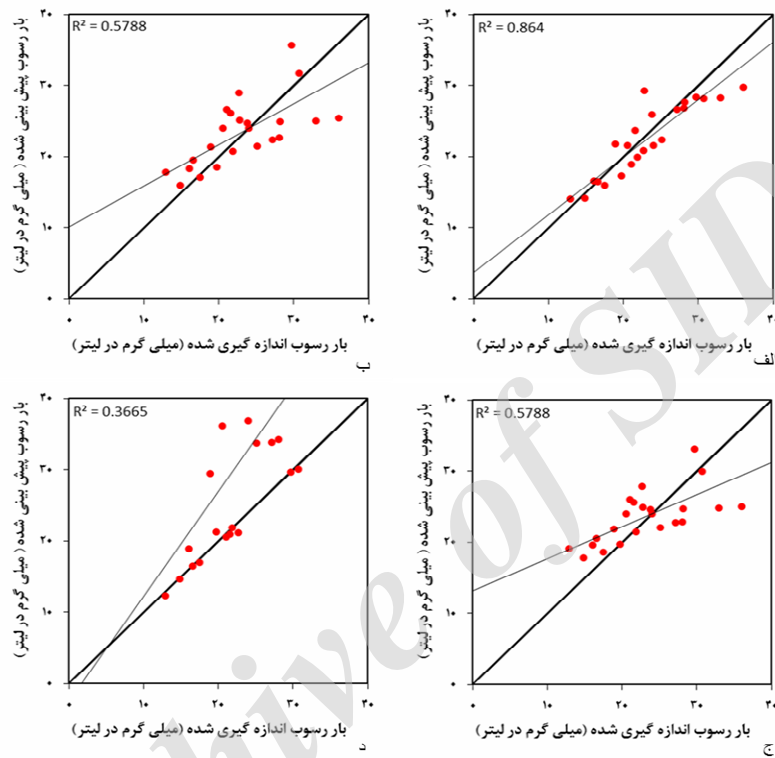
همچنین در بررسی مدل‌ها در داده‌های سالانه نتایج نشان داد که مدل فرآیند گوسی و K نزدیک‌ترین همسایه با مقادیر RMSE به ترتیب ۲/۷ و ۳/۸ دارای بیشترین دقت در پیش‌بینی بار رسوب سالانه می‌باشد. مقادیر پراکنش داده‌های

واقعی و پیش‌بینی شده داده‌های سالانه و نتایج ارزیابی انواع مدل‌ها در شکل‌های ۶ و ۷ نشان داده شده است.



MODEL	FFNN	RBFNN	KNN	LR
RMSE	۸/۶	۵/۴	۸/۳	۴/۵
r	۰/۷۵	۰/۶۲	۰/۸۳	۰/۷۶

شکل (۶) نمودار پراکندگی مقادیر مشاهده شده و پیش‌بینی شده بار رسوب معلق سالانه: توسط مدل‌های الف: شبکه‌ی عصبی مصنوعی پس انتشار خطا، ب: شبکه‌ی عصبی تابع پایه‌ی شعاعی، ج: رگرسیون خطی، د: k نزدیک‌ترین همسایه

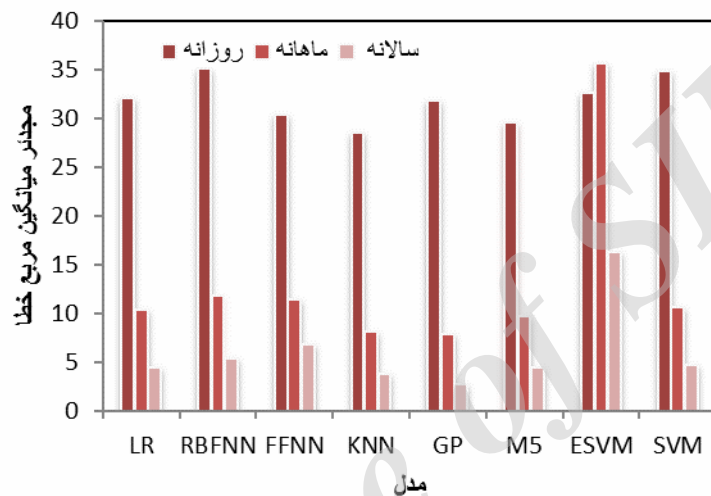


MODEL	GP	M5	SVM	ESVM
RMSE	۷/۲	۴/۵	۴/۶	۱۶/۴
r	۰/۹۳	۰/۷۶	۰/۷۶	۰/۶۰

شکل (۷) نمودار پراکندگی مقادیر مشاهده شده و پیش‌بینی شده بار رسوب معلق ماهانه: توسط مدل‌های الف: فرآیند گوسی، ب: درخت تصمیم M5، ج: ماشین بردار پشتیبان، د: ماشین بردار پشتیبان تکاملی

بررسی مقادیر شاخص‌های ارزیابی در هر سه سری مجموعه داده‌ها بیانگر آنست که میزان دقت مدل‌ها در پیش‌بینی داده‌های سالانه بیش‌تر بوده و در سری

داده‌های روزانه مدل‌ها از دقت کمتری برخوردار بوده‌اند. نمودار مقایسه‌ی شاخص ارزیابی RMSE در شکل (۸) ارائه شده است.



شکل (۸) مقادیر RMSE مدل‌های مختلف بر روی داده‌های رسوب معلق روزانه، ماهانه و سالانه

نتیجه‌گیری

سری‌های زمانی مختلف با توجه به ماهیت‌شان نیازمند مدل‌سازی خاص خود هستند. در این پژوهش به منظور بررسی و ارزیابی کارایی مدل‌ها در پیش‌بینی در سری‌های زمانی مختلف از هشت مدل مختلف استفاده شد. از آنجا که بالا بودن قدرت جریان موجب حمل رسوب بیشتر است لذا در این مطالعه از پارامتر دبی متناظر به منظور مدل‌سازی بار معلق استفاده شد. با اعمال مدل‌های داده‌کاوی بر داده‌های رسوب روزانه مشخص شد که مدل‌های k نزدیک‌ترین همسایه و مدل درختی M5، به ترتیب با مقادیر RMSE ۲۸/۵ و ۲۹/۷ بیشترین دقت را دارند.

در بررسی سری داده‌های ماهانه رفتار مدل متفاوت بوده به طوری که در این سری زمانی مدل‌های فرآیند گوسی و k نزدیک‌ترین همسایه با مقادیر RMSE به ترتیب $۷/۸$ و $۸/۱$ دارای بیشترین دقت بودند. در مجموع دقت مدل‌ها نسبت به سری روزانه بیشتر بوده است. اعمال مدل‌ها بر روی داده‌های سالانه نیز نشان داد که مدل‌های فرآیند گوسی و k نزدیک‌ترین همسایه با مقادیر RMSE $۲/۷$ و $۳/۸$ دارای دقت بالاتری می‌باشند.

در مجموع بررسی مقادیر دقت در سری‌های زمانی مختلف نشان داد که دقت مدل‌ها بر روی داده‌های سالانه نسبت به سایر سری‌های زمانی افزایش یافته است. مطالعات کیاء و عمادی (۱۳۹۲) نیز در تأیید این نتایج است. یکی از دلایل این امر را می‌توان میانگین‌گیری داده‌ها دانست زیرا با میانگین‌گیری تأثیر دبی‌های پائین کمتر شده و در دبی‌های بالا نیز خطای برآورد کمتری دیده می‌شود. بنابراین پیش‌بینی داده‌های روزانه با توجه به بررسی داده‌ها در مقیاس کوچک‌تر و همچنین پراکندگی وسیع داده‌ها نیاز به دقت بالاتری در ساختار مدل خواهد داشت.

منابع

- روشنگر، کیومرث و فریبا پرهیز جوان (۱۳۹۳)، ارزیابی عملکرد سیستم‌های هوش مصنوعی در برآورد بار کل رسوبی رودخانه آجی‌چای، فصلنامه‌ی پژوهشی فضای جغرافیا، شماره‌ی ۴۶، صص ۱۷۳-۱۹۷.
- صنیعی‌آبادی، محمد؛ محمودی، سینا و محدثه طاهرپور (۱۳۹۳)، داده‌کاوی کاربردی (ویراست دوم)، انتشارات نیاز دانش، تهران.
- ظهیری، عبدالرضا و خلیل قربانی (۱۳۹۲)، شبیه‌سازی دبی جریان در مقاطع مرکب به کمک مدل درخت تصمیم M5، نشریه‌ی پژوهش‌های حفاظت آب و خاک، شماره‌ی ۳، صص ۱۱۳-۱۳۲.
- عزمی، محمد و شهاب عراقی‌نژاد (۱۳۹۰)، توسعه روش رگرسیون و k نزدیک‌ترین همسایگی در پیش‌بینی جریان رودخانه، نشریه‌ی آب و فاضلاب، شماره‌ی ۲، صص ۱۱۹-۱۰۸.
- کیا، عیسی و علیرضا عمادی (۱۳۹۲)، مقایسه‌ی روش‌های مختلف رگرسیون آماری در برآورد بار رسوب معلق درازمدت سالانه (مطالعه‌ی موردی: بابل رود)، پژوهشنامه‌ی مدیریت حوضه‌ی آبخیز، شماره‌ی ۸، صص ۱۵-۲۷.
- یوسفی، محسن و ربابه پورشرعیاتی (۱۳۹۳)، برآورد رسوب معلق با استفاده از شبکه‌عصبی و ارزیابی توابع آموزشی (مطالعه‌ی موردی: استان لرستان)، پژوهشنامه‌ی مدیریت حوضه‌ی آبخیز، شماره‌ی ۱۰، صص ۸۵-۹۷.
- Alberg, D., Last, M. and Kandel, A. (2012), **Knowledge Discovery in Data Streams with Regression Tree Methods**, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2(1), PP. 69-78.
- Alp M and Cigizoglu HK, (2007), **Suspended Sediment Load Simulation by two ANN Methods Using Hydrometeorological Data**, Environmental Modeling Software 22: PP. 2-13.

- Bhattacharya, B. and Solomatine, D.P., (2005), **Neural Networks and M5 Model Trees in Modelling Water Level–discharge Relationship**, Neurocomputing, 63, PP. 381-396.
- Cigizoglu, K. (2003). **Estimation and Forecasting of Daily Suspended Sediment Data by Multi-layer Perceptrons**, Advances in Water Resources, 27: PP. 185-195.
- Chang, F.J., Tsai, Y.H., Chen, P.A., Coynel, A. and Vachaud, G., (2015), **Modeling Water Quality in an Urban River Using Hydrological Factors–Data Driven Approaches**, Journal of Environmental Management, 151, PP. 87-96.
- Huang, H.L. and Chang, F.L., (2007), **Evolutionary Support Vector Machine for Automatic Feature Selection and Classification of Microarray Data**, Biosystems, 90(2), PP. 516-528.
- Hornik, K., Stinchcombe, M. and White, H., (1989), **Multilayer Feedforward Networks Are Universal Approximators**, Neural Networks, 2(5), PP. 359-366.
- Kumar, D., Pandey, A., Sharma, N. and Flügel, W.A., (2016), **Daily Suspended Sediment Simulation Using Machine Learning Approach**, Catena, 138, PP. 77-90.
- Lafdani, E.K., Nia, A.M. and Ahmadi, A., (2013), **“Daily Suspended Sediment Load Prediction Using Artificial Neural Networks and Support Vector Machines”**, Journal of Hydrology, 478, pp.50-62.
- Nazari, S., Kuzma, H.A. and Rector III, J.W., (2011), **Predicting Permeability from Well Log Data and Core Measurements Using Support Vector Machines**, In 2011 SEG Annual Meeting, Society of Exploration Geophysicists.
- Ozturk, F., Apaydin, H., & Walling, D.E., (2001), **Suspended Sediment Loads Through Flood Events for Streams of Sakarya Basin**, Turkish Journal of Engineering and Environmental Sciences, 25(6), PP. 643-650.

- Rasouli, K., Hsieh, W.W. and Cannon, A.J., (2012), **Daily Streamflow Forecasting by Machine Learning Methods with Weather and Climate Inputs**, Journal of Hydrology, 414, PP. 284-293.
- Witten, I.H. and Frank, E., (2005). **Data Mining: Practical Machine Learning Tools and Techniques**, Morgan Kaufmann.
- Zounemat-Kermani, M., Kişi, Ö. Adamowski, J. and Ramezani-Charmahineh, A., (2016), **Evaluation of Data Driven Models for River Suspended Sediment Concentration Modeling**, Journal of Hydrology, 535, PP. 457-472.

Archive of SID