

بهبودسازی مدل‌های برآورد بار معلق به کمک پارامترهای زمین ریخت‌شناسی و تکنیک کاهش ویژگی

مریم اسدی^۱، علی فتح‌زاده^{۲*}، روح‌الله تقی‌زاده مهرجردی^۳

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

۲. دانشیار، دانشکده کشاورزی و منابع طبیعی دانشگاه اردکان

۳. استادیار، دانشکده کشاورزی و منابع طبیعی دانشگاه اردکان

(تاریخ دریافت: ۱۳۹۵/۳/۳۰ - تاریخ بازنگری: ۹۵/۱۰/۹ - تاریخ تصویب: ۹۵/۱۰/۱۸)

چکیده

برآورد بار رسوبی رودخانه‌ها از مهم‌ترین چالش‌های مهندسی رودخانه محسوب می‌شود. به همین دلیل تاکنون مدل‌های مختلفی با ساختارهای گوناگون جهت برآورد بار رسوبی ارائه گردیده است. در این مطالعه به بررسی کارایی پارامترهای ژئومورفومتری و تکنیک‌های داده‌کاوی به منظور پیش‌بینی بار رسوب معلق در ۶۸ حوزه واقع در دو منطقه‌ی متفاوت از ایران پرداخته شده است. به همین منظور شش مدل شبکه عصبی مصنوعی، رگرسیون خطی، K نزدیک‌ترین همسایه، فرآیند گوسی، ماشین بردار پشتیبان و ماشین بردار پشتیبان تکاملی بر روی دو نوع از داده‌های بار رسوب معلق (رسوب حداقل و متوسط) اجرا گردید. با هدف بهبودسازی مدل‌ها از پارامترهای ژئومورفومتری و دبی رودخانه به‌عنوان داده‌های ورودی به مدل و از تکنیک کاهش ویژگی به منظور کاهش حجم فضای الگوریتم‌ها استفاده شد. نتایج ارزیابی مدل‌ها نشان داد که عملکرد مدل‌ها در داده‌های رسوب متوسط و حداقل متفاوت می‌باشد، به طوری که بهترین روش جهت پیش‌بینی رسوب متوسط رودخانه مدل فرآیند گوسی با ضریب همبستگی ۰/۹۸۸ و میانگین مجذور مربعات ۱۱/۰۰۴ بوده و مناسب‌ترین روش پیش‌بینی رسوب حداقل مدل ماشین بردار تکاملی با ضریب همبستگی ۰/۹۹۶ و میانگین مجذور مربعات ۰/۱۷۱ می‌باشد.

واژه‌های کلیدی: رسوب معلق، پارامترهای کمکی، داده‌کاوی، کاهش ویژگی، مدل رقومی ارتفاع

مقدمه

برآورد رسوب معلق رودخانه‌ها یکی از فرآیندهای مهم مهندسی رودخانه و منابع آبی می‌باشد و نقش مهمی در طراحی و ساخت سازه‌های آبی و مدیریت فرسایش و رسوب در سطح حوزه‌های آبخیز دارد (Rajaei, 2011). فرسایش سطح حوزه و ورود رسوبات به جریان رودخانه کیفیت منابع آبی و عمر مفید سازه‌های هیدرولیکی را کاهش می‌دهد. لذا برآورد و پیش‌بینی بار رسوبی معلق می‌تواند به کنترل کیفیت منابع آبی و افزایش کارایی تأسیسات برق‌آبی کمک شایانی نماید. اندازه‌گیری بار رسوبی مستلزم احداث ایستگاه‌های هیدرومتری و به‌کارگیری نیروی انسانی و صرف هزینه می‌باشد که در بسیاری از مناطق امکان‌پذیر نمی‌باشد. در این شرایط استفاده از روش‌هایی که بتوان به‌طور غیرمستقیم میزان رسوب را برآورد نمود، مدیریت رسوب و منابع آبی را بهبود خواهد داد. تاکنون مدل‌سازی‌های ریاضی مختلفی در زمینه بار رسوب صورت گرفته است

(Verstraeten and Poesen, 2001; Ward *et al.*, 2009; Choi and Lee, 2015; Wang *et al.*, 2015). مدل‌سازی ریاضی نیاز به حجم وسیعی از اطلاعات داشته که جمع‌آوری و ثبت این اطلاعات بسیار پرهزینه و زمان‌بر می‌باشد. یکی دیگر از روش‌های متداول در زمینه پیش‌بینی بار رسوبی معلق استفاده از منحنی‌های سنج رسوب می‌باشد که به برقراری رابطه بین رسوب و دبی جریان می‌پردازد اما درعین حال برآورد رسوب از طریق این روش خطای زیادی داشته و مقادیر پیش‌بینی بیشتر یا کمتر از مقدار واقعی برآورد گردیده است (Rajaei *et al.*, 2009; Asselman, 2000).

فاکتورهای متعددی در حوزه آبخیز وجود دارد که می‌تواند بر روی مقادیر رسوب معلق حوزه تأثیرگذار باشد و با استفاده از آن‌ها و مدل‌های تجربی می‌توان بار رسوبی را برآورد نمود. در مدل‌های تجربی از پارامترهایی همچون زهکشی، توپوگرافی، دبی، بارش و پوشش گیاهی به‌منظور برآورد بار رسوبی استفاده می‌شود. در این راستا تکنیک‌های داده‌کاوی در برقراری ارتباط بین این عوامل و بار رسوبی به‌خوبی مورد استفاده قرار گرفته‌اند (Lafdani *et al.*, 2013; Kisi, 2012). علم داده‌کاوی مجموعه‌ای از علوم مختلف از جمله آمار، هوش

* نویسنده مسئول: fat@ardakan.ac.ir

ویژگی‌های سطح زمین می‌توان از آن‌ها به‌منظور افزایش دقت مدل‌ها در پیش‌بینی بار رسوبی معلق استفاده نمود.

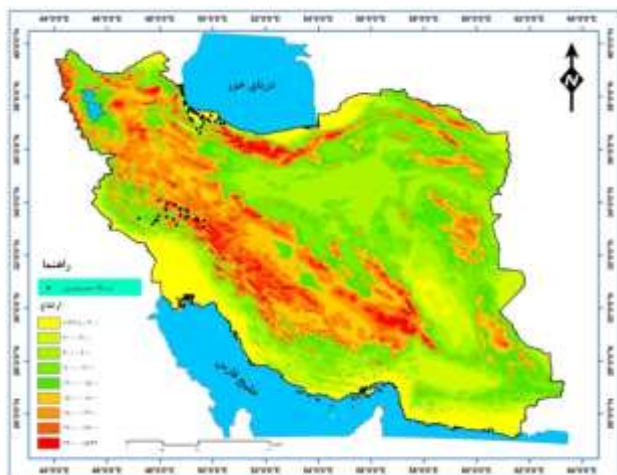
از طرفی یکی از مشکلات مدل‌های داده‌کاوی تهیه منبع داده با ساختار و حجم مناسب می‌باشد. تعداد زیاد پارامترهای ورودی فضای الگوریتم را بزرگ‌تر نموده و باعث پیچیدگی در اجرای الگوریتم می‌گردد. درحالی‌که برخی از همین پارامترهای ورودی به‌کار رفته، نقش چندانی در پیش‌بینی فاکتور موردنظر نداشته و تنها موجب افزایش حجم داده‌های ورودی شده‌اند. به همین دلیل استفاده از تکنیک‌های کاهش ویژگی می‌تواند بدون اتلاف هیچ‌گونه داده‌ای حجم داده‌های ورودی را کاهش دهد (Camdevyren *et al.*, 2005).

در این تحقیق مدل‌های شبکه عصبی مصنوعی، رگرسیون خطی، K نزدیک‌ترین همسایه، فرآیند گوسی، ماشین بردار پشتیبان و ماشین بردار پشتیبان تکاملی انتخاب و با هدف بهینه‌سازی آن‌ها از پارامترهای کمکی حوزه همراه با آنالیز مؤلفه اصلی استفاده گردید تا تأثیر این پارامترها در بهبود مدل‌ها مورد ارزیابی قرار گیرد. همچنین با توجه به تأثیر قدرت جریان در وقوع رسوب حداقل و متوسط، رفتار این مدل‌ها در هرکدام از انواع داده‌های خروجی نیز مورد پایش قرار گرفته است.

مواد و روش‌ها

مناطق مورد مطالعه

مناطق مورد مطالعه ۶۸ حوزه آبخیز واقع در استان‌های گیلان و لرستان ایران بود. این مناطق دارای شرایط مناسب‌تری از نظر طول دوره آماری نسبت به سایر حوزه‌های کشور می‌باشند. حداقل ارتفاع مناطق مورد مطالعه ۸۹ متر و حداکثر آن ۳۷۰۲ متر بود. این دو منطقه در دو اقلیم و شرایط پوشش گیاهی متفاوت واقع شده‌اند. شکل (۱) موقعیت مناطق مورد تحقیق و پراکنش ایستگاه‌های مطالعاتی را نشان می‌دهد.



شکل ۱. موقعیت ایستگاه‌های مورد مطالعه

مصنوعی و یادگیری ماشین می‌باشد که می‌توان با تهیه منبع داده و ارائه صحیح آن به مدل و انجام پردازش‌های مناسب بر روی پایگاه داده‌ها به پیش‌بینی با دقت مناسب رسید. به‌عنوان مثال (Zounemat-Kermani *et al.* 2016) به مدل‌سازی غلظت رسوب معلق بر اساس مدل‌های داده محور پرداختند. در مطالعه آن‌ها مدل‌های شبکه عصبی مصنوعی، SVR و منحنی سنجه رسوب مورد مقایسه قرار گرفتند. نتایج بر اساس معیارهای ارزیابی در این مدل‌ها نشان داد که مدل SVR با تابع RBF دارای عملکرد مناسب‌تری به‌منظور پیش‌بینی بار رسوبی معلق می‌باشد. همچنین (Lafdani *et al.* 2013) از شبکه‌های عصبی مصنوعی و ماشین بردار پشتیبان به‌منظور پیش‌بینی بار رسوبی معلق استفاده نمودند. آن‌ها رواناب و باران را به‌عنوان ورودی مدل و رسوب معلق را به‌عنوان خروجی مدل در نظر گرفتند. نتایج به‌دست‌آمده نشان داد که مدل nu-SVR با استفاده از آزمون گاما دارای عملکرد مناسبی در پیش‌بینی بار رسوبی معلق می‌باشد. Kisi (2012) نیز به مدل‌سازی ارتباط رسوب معلق و دبی با استفاده از حداقل مربعات ماشین بردار پشتیبان (LS-SVM) و شبکه عصبی و منحنی سنجه رسوب پرداخت. جریان روزانه رودخانه و داده غلظت رسوبات معلق در دو ایستگاه بالادست و پایین‌دست حوزه‌ای در منطقه کالیفرنیا به‌عنوان ورودی‌های مدل استفاده شد. نتایج آن‌ها نشان داد که در ایستگاه بالادست مدل LS-SVM بهترین مدل می‌باشد اما در ایستگاه پایین‌دست منحنی سنجه رسوب نتایج بهتری را ارائه داد. استفاده از سایر پارامترهای اقلیمی و هیدرولوژیکی باعث افزایش دقت مدل‌ها در پیش‌بینی بار رسوبی معلق می‌گردد. Zhu *et al.* (2007) به مدل‌سازی جریان رسوب معلق با استفاده از شبکه‌های عصبی مصنوعی و داده‌های متوسط بارش، دما، شدت بارندگی و دبی در رودخانه یانگ تسه چین پرداختند. نتایج آن‌ها نشان داد که ANN هنگامی که متغیرهای مناسب و مؤثر بر جریان رسوب معلق به‌عنوان ورودی استفاده - شود به‌خوبی قادر به مدل‌سازی جریان رسوب معلق ماهانه در مقایسه با مدل‌های رگرسیون خطی چندگانه می‌باشد.

علاوه بر پارامترهای اقلیمی، خصوصیات فیزیوگرافی حوزه نیز در تولید رواناب و رسوب سهم مهمی دارند. باوجود متعدد بودن خصوصیات فیزیوگرافی و سهم آن‌ها در فرایند فرسایش و رسوب‌دهی، تحقیقات محدودی در خصوص بررسی نقش آن‌ها انجام گردیده است. مثلاً (Talebi *et al.* 2016) به بررسی نقش Curvature و انواع آن در میزان فرسایش دامنه‌ها پرداخته‌اند. امروزه به مدد نرم‌افزارهای GIS تعداد قابل‌توجهی از پارامترهای سرزمین را می‌توان استخراج نمود. این پارامترها از مدل رقومی ارتفاعی منطقه مشتق شده‌اند و با توجه به ارتباط آن‌ها با

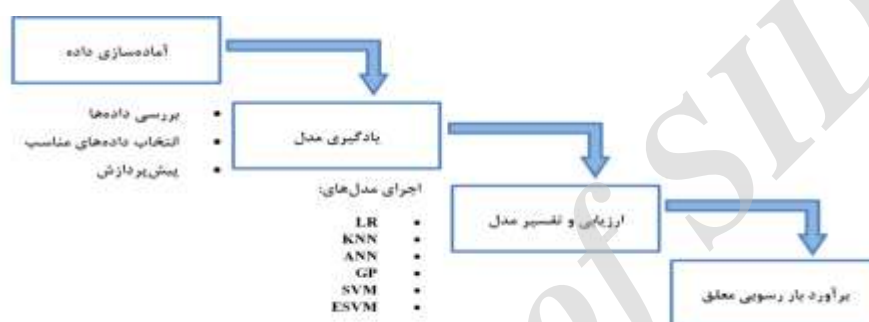
فرآیند مدل‌سازی

جهت مدل‌سازی بار معلق سه مرحله زیر دنبال گردید. الف: آماده‌سازی داده، ب: یادگیری مدل، ج: ارزیابی و تفسیر مدل. در شکل (۲) مراحل کلی پژوهش نمایش داده شده است.

الف) آماده‌سازی داده‌ها

این مرحله شامل فرآیندهایی است که ورودی‌های مناسب را برای مرحله یادگیری مدل فراهم می‌کند. استخراج داده‌ها بایستی از بین منابع موجود صورت گیرد. داده‌های استخراج شده از منابع و نرم‌افزارهای مختلف به صورت متمرکز در یک محل

جمع‌آوری و یک انبار داده تهیه شد. در این مطالعه از داده‌های رسوب ایستگاه‌های واقع در منطقه از سال‌های ۱۹۸۳ تا ۲۰۱۴ میلادی استفاده شد. داده‌های رسوب که به‌عنوان خروجی مدل‌ها تعریف گردید، به دو گروه داده‌های رسوب حداقل حوزه و داده‌های رسوب متوسط حوزه تقسیم شدند و مدل‌ها بر روی هر دو گروه داده‌ها به‌طور مجزا اجرا گردید. مدل رقومی ارتفاعی ۳۰ متر منطقه از سایت USGS دانلود گردید و با استفاده از نرم‌افزار SAGA GIS تعداد ۲۶ پارامترهای ژئومورفومتری از آن استخراج شد (جدول ۱).



شکل ۲. فلوچارت مراحل پژوهش

جدول ۱. پارامترهای ژئومورفومتری مستخرج از DEM

Analytical Hillshading	Flow Directions	Slope
Aspect	General Curvature	Strahler Order
Catchment Area	Longitudinal Curvature	Stream Power Index
Channel Network Base Level	LS Factor	Suspension Load
Convergence Index	MRRTF	Tangential Curvature
Cross-Sectional Curvature	MRVBF	Topographic Wetness Index
Discharge	Plan Curvature	Vertical Distance to Channel Network
Drainage Density	Profile Curvature	Watershed Basins
Flow Accumulation	Relative Slope Position	

تلفات ارائه می‌گردد (Johnson & Wichern, 1982).

ب) یادگیری مدل

این مرحله شامل پیاده‌سازی تکنیک‌های مختلف آماری و داده‌کاوی به‌منظور تعیین الگوی مناسب جهت پیش‌بینی بار رسوب می‌باشد. تکنیک‌های داده‌کاوی به دنبال یافتن رابطه‌ی صحیح بین پارامترهای ورودی و بار رسوبی معلق می‌باشند. مدل‌های مورد استفاده در این تحقیق به شرح زیر می‌باشند:

برای کاهش پیچیدگی تحلیل متغیرها و تفسیر بهتر اطلاعات در مواردی که حجم اطلاعات ورودی زیاد می‌باشد از آنالیز مؤلفه اصلی که از روش‌های آماری چند متغیره است استفاده شد (Camdevyren *et al.*, 2005). در این روش متغیرهای اولیه به‌طور مستقیم استفاده نمی‌شوند بلکه ابتدا به مؤلفه‌هایی تبدیل می‌شوند، سپس این مؤلفه‌ها به‌جای متغیرهای اولیه استفاده می‌شوند. به دلیل استفاده از تمامی متغیرها در تشکیل مؤلفه‌ها، اطلاعات متغیرهای اولیه با کمترین

رگرسیون خطی (LR)

فرآیندهای گوسی، توزیعات گوسی چند متغیره را به بعدیت نامتناهی توسعه می‌دهند. به‌طور رسمی، یک فرآیند گوسی تولید داده‌هایی می‌کند که در طول این طیف قرار دارد به‌طوری‌که هر زیرمجموعه متناهی دامنه یک توزیع گوسی چند متغیره را دنبال می‌کند.

ماشین بردار پشتیبان (SVM)

SVM یک روش محبوب برای حل مسائل طبقه‌بندی، پیش‌بینی و مسائل رگرسیون است (Cortes and Vapnik, 1995). ایده اصلی SVM استفاده از ابر صفحه تفکیک‌کننده خطی است که فاصله بین دو کلاس را برای ایجاد یک دسته‌بندی بیشینه می‌کند. برای ماشین بردار پشتیبان خطی، داده‌های آموزشی متشکل از N جفت $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ می‌باشند که $x_i \in \mathbb{R}^m$ و $y_i \in \{-1, 1\}$, $i=1, \dots, n$ است. فرمول استاندارد ماشین بردار پشتیبان به‌صورت رابطه ۱ می‌باشد:

(رابطه ۱)

$$\min_{\omega, b, \varepsilon} \frac{1}{2} w^T w + c \sum_{i=1}^n \varepsilon_i \text{ subject to } y_i (w^T \phi(x_i) + b) \geq 1 - \varepsilon_i, \quad \varepsilon_i \geq 0, i = 1, \dots, n,$$

که $\omega \in \mathbb{R}^m$ ر وزن نمونه‌های آموزشی است، b ثابت می‌باشد، C یک پارامتر هزینه ارزش‌گذاری شده واقعی و ξ_i یک پارامتر کمکی (متغیر کمکی) می‌باشد. اگر $\phi(x_i) = x_i$ باشد، SVM یک ابر صفحه تفکیک خطی را با حاشیه بیشینه پیدا می‌کند. SVM زمانی که نقشه ϕ ، x_i را بر فضای بعدی بالاتر قرار می‌دهد یک SVM غیرخطی نامیده می‌شود.

ماشین بردار پشتیبان تکاملی (ESVM)

مدل ماشین بردار تکاملی از یک استراتژی تکاملی برای بهینه‌سازی استفاده می‌کند. این مدل یک الگوریتم تکاملی را برای حل مشکل بهینه‌سازی دوگانه ماشین بردار پشتیبان ارائه می‌دهد. پیاده‌سازی این الگوریتم در بسیاری از مجموعه داده‌ها بسیار سریع‌تر و آسان‌تر از ماشین بردار ساده می‌باشد. به‌علاوه، این سیستم قادر به یادگیری با هسته است که نیمه متناهی مثبت نمی‌باشند و می‌توانند برای یادگیری چندمنظوره استفاده شوند که موجب می‌شود تا انتخاب پارامتر C قبل از یادگیری غیرضروری باشد. EVSM بهترین استفاده را از IGA در بهینه‌سازی پارامترهای سیستم با طراحی مدل کروموزومی GA کارآمد دارد و همچنین به‌عنوان یک عملگر هوشمند می‌باشد. ESVM با تجزیه و تحلیل فرکانس از ویژگی‌های انتخاب‌شده، یک مجموعه کوچک از ویژگی‌های مؤثر را شناسایی و از این‌رو باعث بهبود دقت طبقه‌بندی می‌گردد (Huang & Chang, 2007). IGA مورد استفاده به‌طور مختصر ارائه می‌شود

رگرسیون یک روش مورداستفاده برای پیش‌بینی عددی است. رگرسیون یک شاخص و شیوه آماری است که هدف آن تعیین میزان قدرت رابطه بین یک متغیر وابسته (صفت برجسب) و یک سری از متغیرهای در حال تغییر موسوم به متغیرهای مستقل (صفت منظم) می‌باشد. هدف رگرسیون خطی، مدل‌سازی رابطه بین یک متغیر اسکالر و یک یا چند متغیر توضیحی با برازش یک معادله خطی با داده‌های مشاهده‌شده می‌باشد.

شبکه‌های عصبی مصنوعی (ANN)

شبکه‌های عصبی مصنوعی که یک سیستم موازی می‌باشد، متشکل از چندین عنصر پردازشی هستند که توسط لینک‌ها یا حلقه‌های وزن به هم متصل می‌شوند. شبکه‌های عصبی مصنوعی (ANN) دارای یک یا چندلایه پنهان هستند. شبکه‌های عصبی مصنوعی انتشار پس‌خور و پیش‌خور (FFBP)، به‌مراتب از محبوب‌ترین شبکه‌های عصبی مصنوعی می‌باشند (Haykin, 1998). این شبکه متشکل از لایه‌های عناصر پردازشی موازی موسوم به نورون است. هر لایه به‌طور کامل به لایه قبلی با اوزان ارتباطی متصل شده است. مقادیر وزنی که ابتدا داده شده‌اند در هر تکرار در طی فرایند واسنجی (آموزش) تغییر می‌کنند که خروجی‌های برآورد شده را با خروجی‌های معین و معلوم مقایسه می‌کنند و هرگونه خطا برای تعیین تعدیلات وزنی مناسب که به‌منظور کمینه‌سازی خطا لازم هستند را به‌صورت پسین انتشار می‌کنند (Kisi, 2010).

نزدیک‌ترین همسایه K (KNN)

الگوریتم K نزدیک‌ترین همسایه، ساده‌ترین الگوریتم یادگیری ماشینی است. این الگوریتم بر اساس یادگیری با مقایسه است یعنی با قیاس یک مثال آزمایشی معین با مثال‌های آموزشی که مشابه آن هستند اجرا می‌گردد. مثال‌های آموزش توسط n صفت توصیف می‌شوند. هر مثال بیانگر یک نقطه در فضای n بعدی می‌باشند. به این صورت، همه مثال‌های آموزشی در یک فضای الگوی n بعدی ذخیره می‌شوند.

فرآیند گوسی (GP)

فرآیند گوسی یک روش یادگیری ماشینی غیر پارامتری قوی برای ایجاد مدل‌های احتمال گرایانه جامع از مسائل دنیای واقعی است. یک فرآیند گوسی یک فرآیند تصادفی است که متشکل از مقادیر تصادفی در هر نقطه در یک دامنه زمانی یا مکانی است به‌طوری‌که هر یک از متغیرهای تصادفی دارای توزیع نرمال می‌باشند. به‌علاوه، هر مجموعه متناهی از این متغیرهای تصادفی دارای توزیع نرمال چند متغیره است.

است.

همچنین آمار موجود در جدول حاکی از تغییرات وسیع در مساحت حوزه‌های مطالعاتی است. همان‌طور که از اطلاعات آماری جدول مشخص است کوچک‌ترین حوزه دارای مساحت ۲/۶۳ کیلومترمربع و حداکثر مساحت حوزه ۷۷۳۳/۱۲ کیلومترمربع می‌باشد. گسترده بودن دامنه تغییرات در این دو پارامتر موجب گستردگی تغییرات در سایر پارامترهای حوزه‌های مطالعاتی گردیده است.

انتخاب نوع تابع کرنل

در اجرای مدل ماشین بردار پشتیبان با انتخاب نوع تابع کرنل می‌توان دقت مدل را افزایش یا کاهش داد. در واقع انتخاب نوع تابع کرنل یکی از راه‌های ایجاد مدل ماشین بردار پشتیبان کارآمد می‌باشد.

نتایج انواع مختلف کرنل در مدل‌های ماشین بردار پشتیبان و ماشین بردار پشتیبان تکامل‌یافته، اجرا شده بر روی متوسط رسوب حوزه در جدول (۳) و (۴) ارائه شده است. نتایج اعمال تابع کرنل‌های مختلف مدل‌های ماشین بردار پشتیبان بر روی رسوب متوسط نشان داد که تابع کرنل epachnenikov با مقدار $RMSE=۶۵/۶۱۲$ و $r=۰/۷۴$ بهترین نوع تابع کرنل می‌باشد.

در مدل ماشین بردار پشتیبان تکاملی نیز با انتخاب نوع تابع کرنل مقادیر پیش‌بینی با دقت مختلف به دست آمده آمد. با توجه به مقادیر $RMSE$ و r تابع کرنل anova با مقدار $۲۴/۹۵$ و $RMSE=۰/۹۵$ و $r=۰/۹۵$ دارای نتایج بهتری در پیش‌بینی بار رسوبی متوسط می‌باشد.

در بخش دوم از مطالعه انواع تابع کرنل در مدل‌های ماشین بردار پشتیبان و ماشین بردار پشتیبان تکاملی بر روی دسته دوم داده‌ها (حداقل رسوب) مورد بررسی قرار گرفت (جدول ۵ و ۶).

نتایج تابع کرنل در مدل ماشین بردار پشتیبان نشان داد که بر روی رسوب حداقل حوزه نیز تابع کرنل epachnenikov با مقادیر $RMSE=۰/۹۹$ و $r=۰/۸۲$ دارای عملکرد مناسب‌تری نسبت به سایر تابع کرنل‌ها می‌باشد.

نتایج اعمال انواع تابع کرنل در مدل ماشین بردار پشتیبان تکاملی بر روی رسوب حداقل در جدول ۶ نشان می‌دهد که دو تابع کرنل radeial و epachnenikov با مقادیر مشابه $RMSE=۰/۱۷$ و $r=۰/۹۹$ دارای بهترین عملکرد در پیش‌بینی بار رسوبی حداقل می‌باشند.

در حالی که قابلیت طرح‌های آزمایشی متعامد و برتری مقاطع هوشمند را می‌توان در مطالعات دیگر یافت (Ho et al., 2004).

ج) ارزیابی مدل

به منظور ارزیابی مدل از روش ارزیابی K_ Fiod Cross_Validation (Leaving one Out) که در آن مقدار K برابر N استفاده شد. در این روش هر بار یکی از داده‌ها به عنوان داد آزمایشی از داده‌ها جدا شده و فرآیند یادگیری با مابقی داده‌ها (N-1) انجام می‌شود. از فواید این شیوه ارزیابی این است که هر بار یکی از داده‌ها به منظور ارزیابی مورد استفاده قرار می‌گیرد و با توجه به دخالت مجموعه داده‌ها در فرآیند ارزیابی نتایج قابل اعتماد خواهد بود و عیب این روش کندی آن به دلیل تکرار زیاد می‌باشد (Sani Abade et al., 2014).

به منظور ارزیابی نتایج الگوریتم‌های اعمال شده بر روی داده‌ها از معیارهای ارزیابی Root mean squared error (RMSE)، Correlation coefficient (r)، Nash-Sutcliffe (NS) و Mean relative error (MRE) (Najafi et al. 2009) استفاده شد (روابط ۲ الی ۵).

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (o_i - p)^2}{n}} \quad (\text{رابطه ۲})$$

$$NS = 1 - \frac{\sum_{i=1}^n (o_i - p)^2}{\sum_{i=1}^n (o_i - \bar{o})^2} \quad (\text{رابطه ۳})$$

$$r = \frac{\sum_{i=1}^n (p_i - \bar{p})(o_i - \bar{o})}{\sqrt{\sum_{i=1}^n (p_i - \bar{p})^2 \sum_{i=1}^n (o_i - \bar{o})^2}} \quad (\text{رابطه ۴})$$

$$MRE(\%) = \frac{1}{n} \sum_{i=1}^n \left| 100 * \frac{(o_i - p_i)}{o_i} \right| \quad (\text{رابطه ۵})$$

که در این معادلات o مقادیر مشاهده‌ای، p داده‌های پیش‌بینی شده، \bar{p} میانگین داده‌های پیش‌بینی شده، \bar{o} میانگین داده‌های مشاهده شده، n تعداد داده‌ها می‌باشد.

نتایج

خلاصه اطلاعات آماری

پس از استخراج پارامترها و مقادیرشان برای هر یک از زیر حوزه‌ها، اطلاعات آماری آن‌ها بررسی شد. خلاصه آماری داده‌ها در جدول (۲) ارائه شده است. نتایج ارائه شده نشان داد که دامنه تغییرات داده‌ها به دلیل استفاده از ۶۹ حوزه‌ی آبخیز در شرایط متفاوت زیاد بوده است، به طوری که از بین ۶۹ حوزه آبخیز، حداقل ارتفاع حوزه‌های مطالعاتی ۱۵/۹۲ متر (واقع در استان گیلان) و حداکثر ارتفاع ۲۷۸۵/۸۵ متر (واقع در استان لرستان) می‌باشد که در شکل (۱) دو حوزه مطالعاتی نشان داده شده

جدول ۲. اطلاعات آماری پارامترها و داده‌های رسوب و دبی

ضریب تغییرات	انحراف از معیار	میانگین	حداکثر	حداقل	Parameter
۱/۵۷	۸۱۸/۲۴	۱۲۸۷/۷۴	۲۷۸۵/۸۵	-۱۵/۹۲	Dem
۱/۳۱	۱۴۰۷/۸۴	۱۸۵۱/۱۳	۵۷۵۲/۸	۱/۰۵	Watershed Basins
۶/۲۳	۰/۱۵	۰/۲۹	۰/۵۱	۰/۰۱	Slope
۱/۹۵	۰/۳۴	۲/۹۸	۳/۸	۱/۵	Aspect
۸/۷۵	۰/۳۷	۰/۰۴	۳/۰۶	-۰/۰۶	Convergence Index
۰/۱۱	۲۰۰۸۵۱/۱	۳۲۲۱۴۹	۷۶۷۵۳۰/۲	۱۷۶۰۴/۳	Catchment Area
۱/۶	۷۹۸/۲۵	۱۱۶۷/۲۸	۲۵۵۳/۰۳	-۱۹/۸۳	Channel Network Base Level
۱/۴۶	۶۲/۹۶	۱۲۱/۲۳	۳۵۸/۹	۱/۱۸	Vertical Distance to Channel Network
۱/۹۳	۰/۱۱	۰/۴۵	۰/۷۲	۰/۱۱	Relative Slope Position
۳/۸۹	۱/۸۱	۱/۰۳	۶/۶	۰/۰۱	MRVBF
۰/۵۷	۱/۴۹	۰/۷۳	۵/۱۶	۰/۰۱	MRRTF
۱۹/۱۹	۰/۰۸	۱/۴۷	۱/۶۱	۱/۰۸	Strahler Order
-۱۲۰/۵۶	۰/۰۰۱	-۹/۱۳	۰/۰۰۸	-۰/۰۰۰۴	General Curvature
-۱/۰۸	۰/۰۰۰۱	-۰/۰۰۰۱	۰/۰۰۰۸	-۰/۰۰۰۳	Profile Curvature
۰/۱۲	۰/۵۷	۰/۰۷	۴/۷	-۰/۰۱	Plan Curvature
۲/۷۲	۰/۰۰۰۳	۰/۰۰۰۱	۰/۰۰۲۸	-۰/۰۰۰۰۵	Tangential Curvature
-۱/۷۷	۰/۰۰۰۴	-۰/۰۰۰۲	۰/۰۰۲	-۰/۰۰۰۰۵	Longitudinal Curvature
۳/۲۸	۰/۰۰۰۷	۰/۰۰۰۲	۰/۰۰۵	-۰/۰۰۰۰۱	Cross-Sectional Curvature
۶/۵۱	۰/۴۹	۳/۱۸	۴/۴	۱/۶۸	Flow Directions
۱/۷	۳/۳۶	۵/۷۲	۱۰/۹۸	۰/۰۸	LS Factor
۲/۱۴	۷۴۳۹/۵۷	۱۵۹۰۵/۱	۳۵۱۷۰/۲۵	۳۴۵۰/۸	Flow Accumulation
۱/۴	۱۸۱۴۱/۸	۲۵۴۸۱/۹	۶۵۸۸۷/۰۸	۲۱۴/۶۴	Stream Power Index
۵/۷۸	۱/۲۲	۷/۰۵	۱۱/۲۸	۵/۴	Topographic Wetness Index
۱/۷۹	۶۵/۰۲	۱۱۶/۶	۲۹۵/۹۷	۲۷/۵۲	رسوب
۱/۰۳	۲۸/۲۴	۲۷/۲۹	۱۵۳/۵۴	۰/۵۸	دبی
۲/۰۰	۶۳۵/۷۷	۳۱۶/۳۵	۲۷۷۷/۳۷	۱/۲۱	طول آبراهه
۲/۴۳	۱۵۵۵/۳۵	۶۳۸/۷۱	۷۷۳۳/۱۲	۲/۶۳	مساحت حوزه
۰/۴۲	۰/۳۳	۰/۷۸	۱/۷۳	۰/۰۴	تراکم زهکشی

جدول ۳. نتایج انواع تابع کرنل در مدل ماشین بردار پشتیبان اجرا شده بر روی رسوب متوسط

kernel type	RMSE	NSE	MRE	r
dot	۶۵/۱۵	-۰/۰۱۸	۵۲/۷۲	۰/۵۱
radial	۶۵/۵۴	-۰/۰۳۱	۵۱/۳۵	۰/۷۰
polynomial	۶۵/۶۷	-۰/۰۳۵	۵۲/۱۱	۰/۳۷
neural	۶۲/۸۲	۰/۰۵۲	۵۸/۷۱	۰/۲۷
anova	۶۵/۴۳	۰/۰۲۷	۵۲/۷۵	۰/۵۸
epachnenikov	۶۵/۶۱	۰/۰۵۲	۵۱/۳۲	۰/۷۴
multiquadric	۶۴/۵۴	۲e-۸	۶۱/۸۸	۰/۰

جدول ۴. نتایج انواع تابع کرنل در مدل ماشین بردار پشتیبان تکاملی اجرا شده بر روی رسوب متوسط

kernel type	RMSE	NSE	MRE	r
dot	۸۵۸/۴۹	-۱۷۵/۹۲	۷۵۵/۱۵	۰/۲۰
radeial	۲۵/۲۳	۰/۸۴	۱۸/۰۳	۰/۹۵
polynomial	۲۶۵۰/۸	--۱۶۸۵/۸۵	۸۱۳/۰۹	۰/۱۹
anova	۲۴/۹۵	۰/۸۵	۱۷/۸۰	۰/۹۵
epachnenikov	۲۵/۲۳	۰/۸۴	۱۸/۰۳	۰/۹۵

جدول ۵. نتایج انواع تابع کرنل در مدل ماشین بردار پشتیبان اجرا شده بر روی رسوب حداقل

kernel type	RMSE	NSE	MRE	r
dot	۱/۷۴۷	۰/۰۰۵	۱۳۶/۹۹	۰/۲۷۲
radeial	۱/۰۲۸	۰/۵	۱۵۳/۲۹	۰/۷۹۳
polynomial	۱/۳۷۵	۰/۱۳	۱۱۹۲/۰۳	۰/۴۲۸
neural	۸/۳۵۹	-۳۰/۵۱	۹۰۲۲/۲۲	۰/۰
anova	۱/۲۶۸	۰/۲۶	۶۷۴/۷۲	۰/۵۸۸
epachnenikov	۰/۹۹۵	۰/۵۴	۱۵۱/۷۳	۰/۸۲۱
multiquadric	۱/۴۷۸	۰/۰	۲۲۳۸/۷۹	۰/۰

جدول ۶. نتایج انواع تابع کرنل در مدل ماشین بردار پشتیبان تکاملی اجرا شده بر روی رسوب حداقل

kernel type	RMSE	NSE	MRE	r
dot	۲/۷۵	-۲۱/۴۶	۲۵۳۲/۳۹	۰/۰۳۰
radeial	۰/۱۷۱	۰/۹۸	۳۷/۱۰	۰/۹۹۶
polynomial	۲۰/۳۱	-۱۸۷/۸۶	۱۳۷۶۱/۸۵	۰/۰۸۰
anova	۱۶/۳۹۳	۰/۸۸	۳۶/۹۹	۰/۵۲۶
Epachnenikov	۰/۱۷۱	۰/۹۸	۳۷/۱۰	۰/۹۹۶

می‌باشد.

با اعمال مدل‌ها بر روی سری دوم داده‌ها (رسوب حداقل) حوزه نتایج نشان داد که مدل ESVM بهترین مدل جهت پیش‌بینی رسوب حداقل می‌باشد (جدول ۸).

جدول ۷. نتایج ارزیابی مدل‌های اجرا شده بر روی رسوب متوسط

Model	RMSE	NES	MRE	r
ANN	۴۴/۵۶	۰/۳۹	۴۴/۸۲	۰/۸۳
KNN	۴۹/۶۸	۰/۴۰	۴۴/۳۹	۰/۱۶۳
LR	۵۳/۵۲	۰/۳۱	۴۴/۱۹	۰/۵۵
GP_S	۱۱/۰۰	۰/۹۷	۸/۲۵	۰/۹۸
SVM	۶۵/۶۱	۰/۰۵۲	۵۸/۶۷	۰/۷۴
ESVM	۲۴/۹۵	۰/۸۵	۱۹/۵۶	۱/۹۵

انتخاب نوع مدل

با انتخاب بهترین نوع تابع کرنل برای مدل‌های ماشین بردار پشتیبان معمولی و ماشین بردار پشتیبان تکاملی این مدل‌ها همراه با مدل‌های شبکه عصبی مصنوعی، K نزدیک‌ترین همسایه، رگرسیون خطی و مدل گوسی بر روی داده‌های متوسط رسوب اجرا گردید. نتایج نشان داده است که مدل Gaussian Process بهترین مدل جهت برآورد بار رسوب با کمک پارامترهای ژئومورفومتری می‌باشد (جدول ۷).

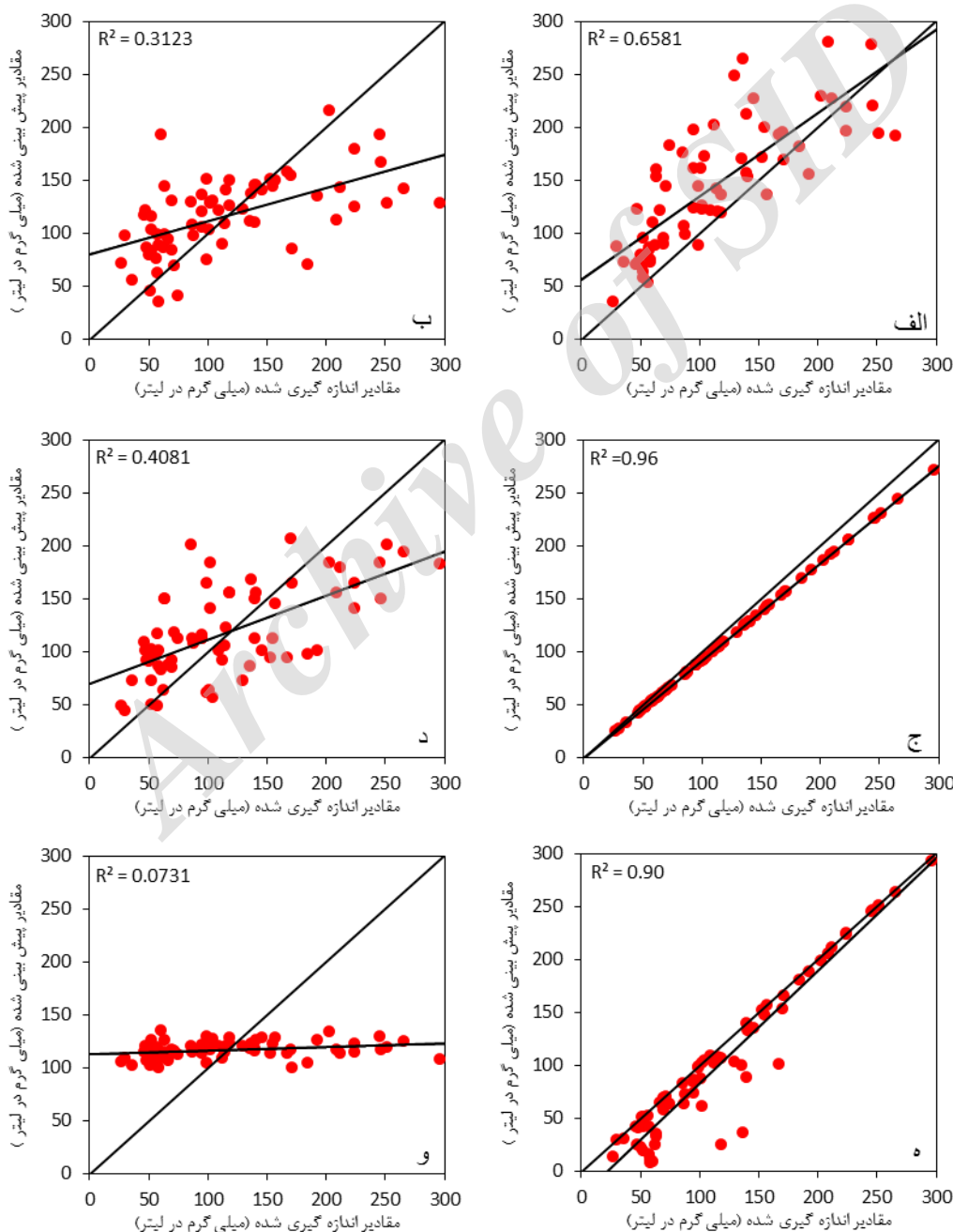
نتایج ارائه شده در جدول (۷) نشان می‌دهد که مدل Gaussian Process با مقدار $RMSE = 11$ و $r = 0.98$ بهترین مدل برای پیش‌بینی بار رسوبی معلق متوسط می‌باشد. مدل ماشین بردار پشتیبان تکاملی نیز بعد از مدل Gaussian Process با مقدار $RMSE = 24.95$ و ضریب $r = 0.95$ دارای عملکرد نسبتاً مناسبی جهت پیش‌بینی بار رسوبی معلق متوسط

جدول ۸. نتایج ارزیابی مدل‌های اجرا شده بر روی رسوب حداقل

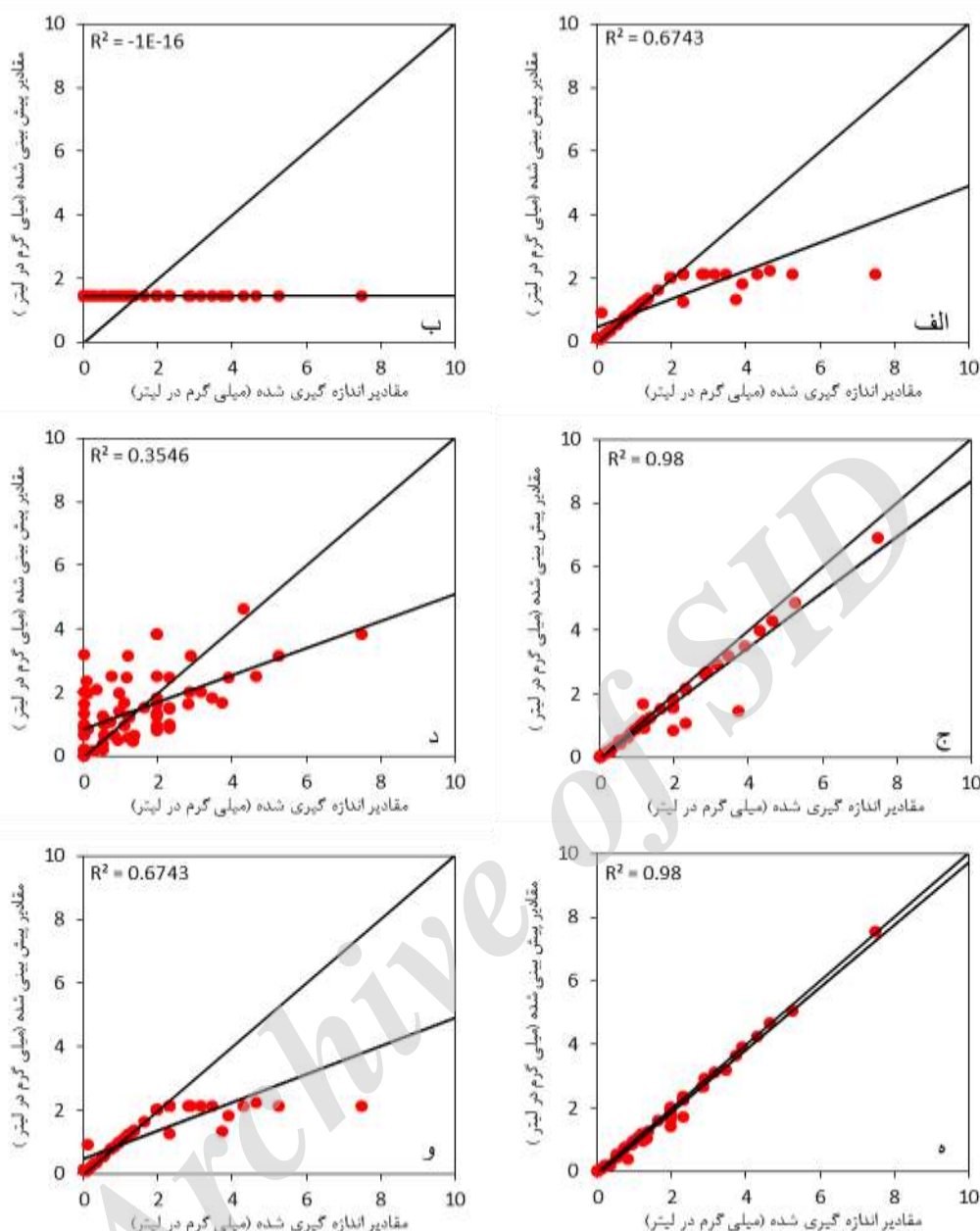
Model	RMSE	NSE	MRE	r
ANN	۰/۸۷	۰/۶۵	۹۶۲/۴۰	۰/۸۲
KNN	۱/۲۰	۰/۳۴	۱۳۴۰/۹۴	۰/۵۹
LR	۰/۵۳	۰/۰۰۶	۲۲۳۸/۷۹	۰/۹۳
GP _s	۰/۱۹	۰/۹۸۲	۱۵/۰۸	۰/۹۹
SVM	۰/۹۹	۰/۵۴	۱۵۸/۳۶	۰/۸۲
ESVM	۰/۱۷	۰/۹۸۶	۱۵/۸۹	۰/۹۹

رسوب حداقل نتایج متفاوتی را نسبت به رسوب متوسط به دنبال دارد. به طوری که در مورد رسوب حداقل مدل ESVM با مقدار $RMSE = ۰/۱۷$ و $r = ۰/۹۹$ دارای عملکرد بهتری نسبت به سایر مدل‌ها می‌باشد. البته اختلاف دو مدل GP_s و ESVM بسیار کم می‌باشد و می‌توان چنین نتیجه گرفت که از میان مدل‌های اجرا شده بر روی هر دو سری داده دو مدل GP_s و مدل ESVM دارای بهترین عملکرد بوده است. پراکنش مقادیر داده‌های واقعی و پیش‌بینی شده در مدل‌های اجرا شده در شکل‌های (۳) و (۴) نشان داده شده است:

نتایج جدول (۸) نشان داد اجرای مدل‌ها بر روی داده‌های



شکل ۳. نمودار پراکندگی مقادیر مشاهده شده و پیش‌بینی شده رسوب متوسط توسط مدل‌های الف: شبکه عصبی مصنوعی، ب: رگرسیون خطی، ج: فرآیند گوسی، د: k نزدیک‌ترین همسایه، ه: ماشین بردار پشتیبان تکاملی و و: ماشین بردار پشتیبان



شکل ۴. نمودار پراکندگی مقادیر مشاهده‌شده و پیش‌بینی‌شده رسوب حداقل توسط مدل‌های الف: شبکه عصبی مصنوعی، ب: رگرسیون خطی، ج: فرآیند گوسی، د: k نزدیک‌ترین همسایه، ه: ماشین بردار پشتیبان تکاملی و و: ماشین بردار پشتیبان

بحث و نتیجه‌گیری

استفاده از تکنیک‌های کاهش ویژگی موجب کاهش حجم فضای الگوریتم و پیچیدگی‌های آن و سهولت در اجرای الگوریتم‌ها می‌شود. در این مطالعه نیز به دلیل زیاد بودن تعداد پارامترها از این تکنیک به منظور کاهش حجم داده‌ها استفاده گردید.

از طرفی نتایج حاصل از ارزیابی مدل‌ها نشان داد که رفتار دو رسوب متوسط و حداقل در شبیه‌سازی‌های صورت گرفته یکسان نبوده و در نتیجه نوع رسوب نیز در انتخاب مدل بهینه تأثیرگذار می‌باشد. بدین ترتیب که برای داده‌های رسوب متوسط مدل Gaussian Process با ضریب همبستگی ۰/۹۸۸ و

در این تحقیق به کمک برخی پارامترهای ژئومورفومتری به بهینه‌سازی مدل‌های برآورد رسوب پرداخته شد. استفاده از پارامترهای ژئومورفومتری باعث افزایش دقت مدل در پیش‌بینی رسوب معلق حوزه شد. این درحالی‌که است که مدلهایی که داده‌های ورودی آن‌ها صرفاً داده‌های اقلیمی بوده است از دقت کمتری برخوردار بوده‌اند (Cobaner et al., 2009; Liu et al., 2013; Zounemat-Kermani et al., 2016). روش کلاسیک منحنی سنج رسوب نیز صادق است (Zounemat-Kermani et al., 2016; Kisi, 2012; Rajaei, 2011).

به نظر می‌رسد با وارد ساختن سایر پارامترهای مؤثر در فرسایش و رسوب‌دهی حوزه (مثل پوشش گیاهی، کاربری اراضی، تصاویر ماهواره‌ای) افزایش دقت مدل‌ها را می‌توان انتظار داشت. همچنین بررسی کارایی سایر مدل‌های داده‌کاوی و الگوریتم‌های مبتنی بر هوش مصنوعی نیز در ادامه این تحقیق پیشنهاد می‌گردد.

میانگین مجذور مربعات $11/004$ به‌عنوان بهترین روش تعیین شد و برای داده‌های رسوب حداقل، مدل ESVM با ضریب همبستگی $0/996$ و میانگین مجذور مربعات $0/171$ مناسب‌ترین روش معرفی گردید. بدین ترتیب نتایج به‌دست‌آمده نشان داده است که تلفیق تکنیک‌های داده‌کاوی و پارامترهای ژئومورفومتری می‌تواند به نحو چشمگیری باعث افزایش دقت در پیش‌بینی رسوب معلق گردد.

REFERENCES

- Asselman, N.E.M. 2000. Fitting and interpretation of sediment rating curves. *Journal of Hydrology*. 23 (4), 228-248.
- Camdevyren, H. Demyr, N. Kanik, A. and Keskin, S. 2005. Use of principal componentscores in multiple linear regression models for prediction of Chlorophyll-a in reservoirs. *Ecological Modelling*. 181(4), 581-589.
- Choi, S.U. and Lee, J., 2015. Assessment of total sediment load in rivers using lateral distribution 12. Cortes, C., Vapnik, V., 1995. Support-vector network. *Mach. Learn.* 20, 273-297.
- Cobaner, M., Unal, B. and Kisi, O., 2009. Suspended sediment concentration estimation by an adaptive neuro-fuzzy and neural network approaches using hydro-meteorological data. *Journal of hydrology*, 367(1), pp.52-61. *method. Journal of Hydro-environment Research*, 9 (3), pp.381-387.
- Haykin, S., 1998. *Neural Networks – A Comprehensive Foundation*, second ed. Prentice-Hall, Upper Saddle River, NJ, pp. 26-32.
- Ho, S.-Y., Shu, L.-S., Chen, J.-H., 2004. Intelligent evolutionary algorithms for large parameter optimization problems. *IEEE Trans. Evolutionary Comput.* 8 (6), 522-541.
- Huang, H.L. and Chang, F.L., 2007. ESVM: Evolutionary support vector machine for automatic feature selection and classification of microarray data. *Biosystems*, 90(2), pp.516-528.
- Johnson, R. A. and Wichern, D. W. 1982. *Applied multivariate statistical analysis*, 3rd Ed, Prentice-Hall Inc, Englewood Cliffs, USA.
- Kisi, O., 2010. River suspended sediment concentration modelling using a neural differential evolution approach. *J. Hydrol.* 389 (1-2), 227-235
- Kisi, O., 2012. Modeling discharge-suspended sediment relationship using least square support vector machine. *Journal of hydrology*, 456, pp.110-120.
- Liu, Q.J., Shi, Z.H., Fang, N.F., Zhu, H.D. and Ai, L., 2013. Modeling the daily suspended sediment concentration in a hyperconcentrated river on the Loess Plateau, China, using the Wavelet-ANN approach. *Geomorphology*, 186, pp.181-190.
- 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.
- Najafi, G., Ghobadian, B., Tavakoli, T., Buttsworth, D.R., Yusaf, T.F. and Faizollahnejad, M., 2009. Performance and exhaust emissions of a gasoline engine with ethanol blended gasoline fuels using artificial neural network. *Applied Energy*, 86(5), pp.630-639.
- Rajaei, T., 2011. Wavelet and ANN combination model for prediction of daily suspended sediment load in rivers. *Science of the total environment*, 409 (15), pp.2917-2928.
- Rajaei, T., Mirbagheri, S.A., Zounemat-Kermani, M. and Nourani, V., 2009. Daily suspended sediment concentration simulation using ANN and neuro-fuzzy models. *Science of the total environment*, 407(17), pp.4916-4927.
- Sani Abade, M., Mahmoudi, S. and Taherparvar, D. (2014). *Data mining applications (second edition)*, Niaz-e-Danesh Pub. Tehran. (In Farsi).
- Talebi, A., Hajiabolfhasemi, R., Hadian, M.R. and Amanian, N., 2016. Physically-based modeling of sheet erosion (detachment and deposition processes) in complex hillslopes. *Hydrological Processes*. 30(12). pp.1968-1977.
- Verstraeten, G. and Poesen, J., 2001. Factors controlling sediment yield from small intensively cultivated catchments in a temperate humid climate. *Geomorphology*, 40(1), pp.123-144.
- Ward, P.J., van Balen, R.T., Verstraeten, G., Renssen, H. and Vandenberghe, J., 2009. The impact of land use and climate change on late Holocene and future suspended sediment yield of the Meuse catchment. *Geomorphology*, 103(3), pp.389-400.
- Wang, Y.G., Wang, S.S. and Dunlop, J., 2015. Statistical modelling and power analysis for detecting trends in total suspended sediment loads. *Journal of Hydrology*, 520, pp.439-447.
- Zhu, Y.M., Lu, X.X. and Zhou, Y., 2007. Suspended sediment flux modeling with artificial neural network: an example of the Longchuanjiang River in the Upper Yangtze Catchment, China. *Geomorphology*, 84(1), pp.111-125.
- Zounemat-Kermani, M., Kisi, 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.