



ص ۳۴۵-۳۵۸

تاریخ دریافت: ۱۳۹۴/۰۶/۲۳

تاریخ تصویب: ۱۳۹۵/۱۲/۱۵

مقایسه کارایی روش‌های رگرسیون بردار پشتیبان و k-نزدیک‌ترین همسایگی در برآورد میزان بار رسوبی معلق در رودخانه (مطالعه موردی: رودخانه ليقوان چای)

❖ **علی رضازاده جودی***: کارشناسی ارشد عمران-آب، باشگاه پژوهشگران جوان و نخبگان، واحد مراغه، دانشگاه آزاد اسلامی، مراغه، ایران.

❖ **محمد تقی ستاری***: استادیار، مهندسی منابع آب، گروه مهندسی آب، دانشکده کشاورزی، دانشگاه تبریز، ایران.

چکیده

برآورد بار رسوبی معلق رودخانه‌ها با توجه به خسارات ناشی از عدم توجه و لحاظ کردن آن، یکی از مهم‌ترین و اساسی‌ترین چالش‌های مطالعات انتقال رسوب و مهندسی رودخانه می‌باشد. با توجه به اهمیت و نقش رسوب در طراحی و نگهداری سازه‌های هیدرولیکی همچون سدها و همچنین برنامه‌ریزی جهت استفاده بهینه از منابع آبی در پایین‌دست رودخانه‌ها و حفظ منابع مغذی بالادست آن‌ها، همواره تلاش‌های بسیاری در زمینه تخمین میزان بار رسوبی معلق رودخانه‌ها انجام گرفته و روش‌های متعددی در این زمینه توسعه یافته است. اما با توجه به هزینه‌بر بودن اکثر روش‌ها و یا عدم دقت کافی در اکثر روش‌های تجربی مرسوم، نیاز به روش نوینی که بتواند بار رسوبی معلق رودخانه را با بیشترین دقت ممکن تخمین زند، امری ضروری به نظر می‌رسد. در این مطالعه میزان بار رسوبی معلق رودخانه ليقوان چای توسط روش‌های رگرسیون بردار پشتیبان و k-نزدیک‌ترین همسایگی برآورد گردیدند. نتایج نشان‌دهنده عملکرد مناسب هر دو روش داده‌کاوی بررسی شده در این تحقیق می‌باشد. از میان روش‌های بررسی شده در این تحقیق، روش رگرسیون بردار پشتیبان میزان بار رسوبی معلق رودخانه ليقوان چای را با ارائه مقادیر ضریب همبستگی برابر با ۰/۹۵۹ و ریشه میانگین مربعات خطا برابر با ۴۳/۵۴۷ (تن در روز) با دقت بیشتری نسبت به روش k-نزدیک‌ترین همسایگی پیش‌بینی کرد.

کلید واژگان: بار رسوبی معلق، داده‌کاوی، رگرسیون بردار پشتیبان، ليقوان چای، k-نزدیک‌ترین همسایگی.

۱. مقدمه

با توجه به اینکه مشکل کمیت و کیفیت آب همواره یکی از چالش‌های اساسی پیش‌روی جوامع انسانی می‌باشد، همواره حفظ و نگهداری منابع موجود و در دسترس از اولویت بالایی برخوردار بوده است. برآورد بار رسوبی رودخانه‌ها با توجه به خسارات ناشی از عدم توجه و لحاظ کردن آن، یکی از مهم‌ترین و اساسی‌ترین چالش‌های مطالعات انتقال رسوب و مهندسی رودخانه می‌باشد. بار کل رسوب عبارت است از حاصل جمع بار بستر و بار معلق و یا حاصل جمع بار مواد بستر و بار مواد آبرفتی، موقعی که رسوبات در رودخانه کم است و در رودخانه‌های کم عمق، کل بار رسوب تقریباً برابر با بار بستر خواهد بود. برعکس در رودخانه‌های عمیق بار بستر فقط بین ده تا بیست درصد بار کل را تشکیل می‌دهد [۱۸]. تخمین میزان بار رسوب یا میزان انتقال آن برای بسیاری از کاربری‌ها و پروژه‌های منابع آب مانند طراحی سدها و حوضچه‌های ذخیره آب سطحی، انتقال رسوب و آلودگی در رودخانه‌ها و دریاچه‌ها، طراحی کانال‌ها و نگهداری آن‌ها، حیاتی می‌باشد [۲۵]. بار معلق رسوبی رودخانه‌ها معمولاً از طریق اندازه‌گیری مستقیم یا معادلات تجربی تعیین می‌شود. گرچه اندازه‌گیری مستقیم از درجه اعتبار زیادی برخوردار است، ولی بسیار هزینه‌بر بوده و نمی‌تواند برای سطح وسیعی از رودخانه مورد استفاده قرار گیرد. با این حال اکثر معادلات برآوردکننده بار معلق رسوبی، نیازمند داده‌های دبی و رسوب حاصل از اندازه‌گیری‌ها می‌باشند [۱۲]. روابط تجربی متعددی در این زمینه ارائه شده است که در بسیاری از موارد نتایج حاصل از این روابط به دلیل پیچیدگی و عدم شناخت کامل و دقیق مکانیزم انتقال رسوب، با مقادیر اندازه‌گیری شده تطابق نداشته و در مواردی نیز روابط مختلف تفاوت معنی‌داری با یکدیگر دارند [۲۲]. از سوی دیگر روابط موجود تحت شرایط خاصی توسعه یافته است و استفاده از آن‌ها در شرایطی

دیگر نیازمند بررسی و مقایسه نتایج با ارقام اندازه‌گیری شده است. یکی از روش‌های متداول محاسبه بار معلق رودخانه استفاده از منحنی سنج رسوب^۱ (SRC) است. این رابطه به دلیل داشتن اریب^۲ در بیشتر مواقع نتوانسته غلظت رسوب در دبی‌های مختلف جریان را به خوبی نشان دهد، اریب در واقع باعث شده تا مقدار باقی‌مانده‌ها (اختلاف بین مقادیر مشاهده‌ای و محاسباتی) توزیع نرمالی نداشته باشد [۷]. اریب منحنی سنج از دو عامل منشأ می‌گیرد. عامل اول به علت تغییر شکل معادله از حالت لگاریتمی به حالت مدل طبیعی رگرسیون خطی است، که در اصل مربوط به ذات منحنی سنج می‌باشد. عامل دوم اریب ناشی از برون‌یابی برای برآورد رسوب دبی‌های بالاست [۵]. تمامی این عوامل سبب شده است تا در دهه اخیر تلاش‌ها برای پیدا کردن روشی که محدودیت‌ها و مشکلات موجود در روابط و روش‌های کلاسیک متداول را نداشته باشد، بیشتر شود. ازجمله این روش‌ها می‌توان به روش‌های موسوم به محاسبات نرم اشاره کرد. اساس این روش‌ها پیدا کردن الگوها و روابط پنهان بین داده‌ها می‌باشد. از جمله تحقیقات صورت گرفته در این زمینه می‌توان به تحقیقی [۲۶] در زمینه استفاده از شبکه عصبی مصنوعی برای مدل‌سازی و برآورد مقدار رسوب معلق در رودخانه‌ای در چین اشاره کرد که کارایی بسیار مناسب این روش را برای این موضوع اعلام کردند. در پژوهشی [۱۳] نشان داده شد که در تخمین میزان بار رسوب معلق، استفاده هم‌زمان از دو روش شبکه عصبی مصنوعی و موجک، تطابق بسیار مناسبی بین داده‌های مشاهداتی و محاسباتی فراهم می‌کند. در مطالعه دیگری [۳] مقایسه‌ای بین عملکرد دو فرمول برآورد بار معلق در رودخانه انجام داده شد و نتایج نشان داد که از بین روش‌های مذکور روش منحنی سنج رسوب دارای دقت کمتری می‌باشد. همچنین در پژوهش دیگری [۱۱] با استفاده از مدل‌سازی چند متغیره اقدام

1- Suspended Rating Curve

2- Bias

هرمزگان را با استفاده از پنج روش برآورد کردند، نتایج نشان داد که معادله اصلاح شده انشتین پاسخ واقعی تری را ارائه کرده است. در تحقیق دیگری [۴] بار رسوب رودخانه با استفاده از شبکه عصبی مصنوعی مدل‌سازی گشته و نتایج نشان داد که، مدل‌های ارائه شده توسط شبکه عصبی مصنوعی در مقایسه با روابط تجربی نتایج دقیق‌تری ارائه کردند. در پژوهش دیگری [۸] رسوبات معلق رودخانه قطورچای پل یزدکان با استفاده از ترکیب منحنی سنج رسوب و شبکه عصبی مصنوعی پیش‌بینی گردید. نتایج نشان داد که، کاربرد ترکیبی شبکه عصبی مصنوعی و منحنی سنج رسوب باعث بالا رفتن دقت مدل در مقایسه با کاربرد مجزای هر کدام از روش‌های مذکور می‌شود. در مطالعه دیگری [۲۱] با استفاده از شبکه عصبی مصنوعی و خوشه‌بندی داده‌ها به روش نگاشت خودسازمان‌ده به برآورد غلظت رسوب معلق روزانه در رودخانه کرج پرداخته شد و نتایج نشان داد که استفاده از متغیرهای بارش و دمای روزانه، دبی جریان و تفکیک داده‌ها نقش مهمی در افزایش دقت برآورد رسوب رودخانه داشته است. در مطالعه دیگری [۱۴] با استفاده از مدل تبرید تدریجی عصبی (NDE) و مدل‌های ANFIS و RBF به تخمین بار معلق رسوبی پرداختند و دریافتند که مدل تبرید تدریجی قادر به ارائه نتایج دقیق‌تری می‌باشد. در پژوهش دیگری [۱۷] عملکرد روش‌های مدل درختی M5 و رگرسیون بردار پشتیبان در برآورد بار رسوبی معلق در رودخانه اهرچای مورد ارزیابی قرار گرفت. نتایج نشان داد که هر دو این روش‌ها دارای دقت بالا و عملکرد مناسبی هستند منتهی روش مدل درختی M5 با ارائه روابط خطی ساده کاربردی‌تر می‌باشد. همچنان‌که مشاهده می‌شود تحقیقات بسیاری در مورد کارایی و عملکرد شبکه‌های عصبی مصنوعی در زمینه برآورد میزان بار معلق رودخانه‌ها انجام پذیرفته است اما مطالعات چندانی در زمینه کاربرد سایر روش‌های محاسبات نرم از جمله مدل‌های رگرسیون بردار پشتیبان

به تخمین بار معلق در بالادست حوزه آبخیز گشته و نتایج نشان داد که مدل درختی M5 قابلیت خوبی برای این مسئله دارد. در تحقیق دیگری [۹] با استفاده از شبکه‌های عصبی مصنوعی به تخمین میزان بار معلق رودخانه بابل‌رود پرداختند و مشاهده کردند که بهترین نتیجه مربوط به استفاده از شبکه عصبی پرسپترون چندلایه است و در حالت کلی نتایج شبکه عصبی بهتر از نتایج منحنی سنج رسوب است. در مطالعه دیگری [۶] با استفاده از دو مدل شبکه عصبی مصنوعی و ماشین بردار پشتیبان به پیش‌بینی بار معلق رسوب روزانه رودخانه دویرج پرداختند و نتایج را با مدل‌های رگرسیونی مقایسه نمودند. نتایج نشان داد که شبکه عصبی مصنوعی و پس از آن ماشین بردار پشتیبان دقت عمل بیشتری نسبت به مدل‌های رگرسیونی در پیش‌بینی رسوب دارند. در پژوهشی [۱۰] به معرفی مدل‌های هیبریدی مبتنی بر هوش مصنوعی برای تخمین میزان بار رسوبی معلق پرداخته شد. در مطالعه دیگری [۱۶] پارامترهای کیفی آب شامل میزان هدایت الکتریکی و کل جامدات محلول با استفاده از روش رگرسیون بردار پشتیبان و k-نزدیک‌ترین همسایگی تخمین زده شد که نتایج نشان‌دهنده کارایی بالای این روش‌ها در این زمینه بود. از جمله تحقیقات صورت گرفته در ایران در این زمینه می‌توان به تحقیق [۲] اشاره کرد که ضمن مقایسه شبکه‌های عصبی مصنوعی با روش منحنی سنج رسوب در تخمین رسوب معلق، نشان دادند که این شبکه‌ها توانایی بیشتری در پیش‌بینی رسوب معلق رودخانه بهشت آباد از سرشاخه‌های کارون دارند. در پژوهش دیگری [۲۲] بار رسوب جریان در حوزه آبخیز سمندگان با استفاده از شبکه‌های عصبی مصنوعی و رگرسیون پیش‌بینی شد، که مشاهده گردید شبکه عصبی ژنومورفولوژی در مقایسه با شبکه عصبی غیر ژنومورفولوژی و مدل‌های رگرسیونی ژنومورفولوژی و غیر ژنومورفولوژی دقت بالاتری دارد. در مطالعه دیگری [۱۵] رسوب در رودخانه جگین در استان

دامنه شمالی سهند بین طول‌های ۴۶ درجه و ۲۰ دقیقه و ۳۰ ثانیه تا ۴۶ درجه و ۲۷ دقیقه و ۳۰ ثانیه شرقی و عرض‌های ۳۷ درجه و ۴۲ دقیقه و ۵۵ ثانیه تا ۳۷ درجه و ۴۹ دقیقه و ۳۰ ثانیه شمالی گسترده شده است. ارتفاع بلندترین نقطه آن ۳۶۲۰ متر و پایین‌ترین نقطه آن ۲۱۴۰ متر بالاتر از سطح آب‌های آزاد می‌باشد. به طور کلی حوزه آبریز ليقوان به علت وجود پوشش‌های برفی دایمی در ارتفاعات مختلف آن دارای رژیم برفی است. ذوب برف در جریان دایمی رودخانه‌های آن مؤثر و قابل ملاحظه است. ریزش باران در ماه‌های بهار و تابستان شدید بوده و این امر اثر قابل توجهی در بالا بردن مقدار دبی رودخانه‌ها دارد. موقعیت جغرافیایی حوزه ليقوان در شکل (۱) نشان داده شده است.

و k-نزدیک‌ترین همسایگی در این زمینه صورت نپذیرفته است. هدف از این تحقیق پیدا کردن و معرفی راه حل و روشی جایگزین برای برآورد میزان بار معلق با دقت و کارایی بالا می‌باشد. برای این منظور کارایی و عملکرد روش رگرسیون بردار پشتیبان و k-نزدیک‌ترین همسایگی در برآورد میزان بار معلق رودخانه ليقوان چای واقع در استان آذربایجان شرقی، مورد سنجش و ارزیابی قرار می‌گیرد.

۲. روش شناسی تحقیق

۱،۲. معرفی منطقه مورد مطالعه

حوزه آبریز ليقوان یکی از زیر حوزه‌های دریاچه ارومیه بوده و با وسعتی معادل ۷۶/۱۹ کیلومتر مربع در



شکل ۱. موقعیت جغرافیایی منطقه مورد مطالعه (حوزه آبریز ليقوان)

۲,۲. داده‌های استفاده شده

در این تحقیق از داده‌های مربوط به پارامترهای دبی (مترمکعب در ثانیه)، غلظت بار معلق (میلی‌گرم در لیتر)، رسوب (تن در روز) رودخانه ليقوان چای مربوط به سال‌های آبی ۷۳ تا ۹۳، که در مجموع شامل ۳۱۰

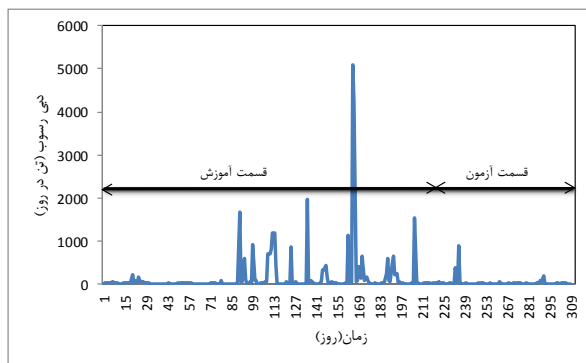
مجموعه داده می‌باشد، استفاده شده است. از این تعداد داده، ۷۵٪ داده‌ها برای قسمت آموزش و ۲۵٪ داده‌ها برای قسمت آزمون مدل در نظر گرفته شد. مشخصات آماری مربوط به داده‌های مورد استفاده در جدول (۱) ارائه گردیده است.

جدول ۱. مشخصات آماری داده‌های مورد استفاده

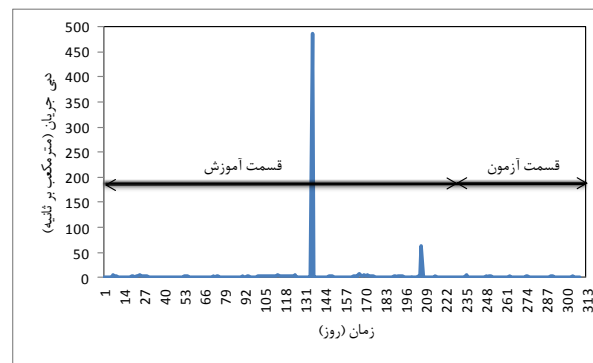
	دبی رسوب (Ton/Day)	غلظت مواد معلق (Mg/litr)	دبی (m ³ /s)
حداقل	۰/۰۲۲	۱/۲۵	۰/۰۰۶
حداکثر	۵۰۸۹/۷	۱۱۳۷۲/۳۱	۴۸۶
میانگین	۱۱۱/۵۱۵	۵۲۲/۰۸۶	۲/۹۱۲
انحراف معیار	۴۴۱/۹۷۱	۱۱۷۰/۳۰۲	۲۷/۷۶۷

استاندارد بودن توسط نرم‌افزار XLSTAT و با روش نرمال استاندارد، ون نیومن، بیشند و پتیت مورد آزمون و بررسی قرار گرفتند. در شکل (۲) و (۳) تغییرات زمانی میزان دبی جریان و دبی رسوبی در رودخانه ليقوان چای ارائه شده است.

با توجه به اینکه، عدم استفاده از داده‌های استاندارد و همگن، تأثیرات بسزایی در میزان دقت و کارایی مدل‌های بدست آمده از روش‌های داده‌کاوی داشته، در نتیجه سبب کاهش قدرت تعمیم‌دهی آن‌ها می‌شود. در تحقیق حاضر، قبل از اقدام به مدل‌سازی میزان دبی بار معلق رودخانه ليقوان چای، تمامی داده‌ها از نظر همگن بودن و



شکل ۳. تغییرات زمانی دبی رسوب



شکل ۲. تغییرات زمانی دبی جریان

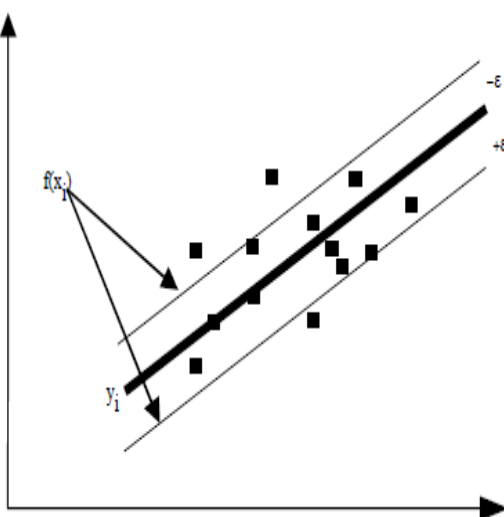
طبقه‌بندی‌کننده بردار پشتیبانی و رگرسیون بردار پشتیبانی. رگرسیون بردار پشتیبان از دقیق‌ترین و

۳,۲. رگرسیون بردار پشتیبان

ماشین‌های بردار پشتیبان شامل دو دسته می‌باشند:

پشتیبان با فرض مقادیر ورودی X_i و مقادیر خروجی Y_i ، هدف یافتن تابعی است که کمترین انحراف (ϵ) را از Y_i ها داشته باشد، ϵ میزان انحراف می‌باشد [۲۴]. در شکل (۴) شمایی از خطی‌سازی به کمک الگوریتم ماشین بردار پشتیبان با در نظر گرفتن انحراف به مقدار ϵ نشان داده شده است. در برخی مواقع، به ساختارهای پیچیده و غیرخطی برای جداسازی داده‌ها نیاز است. در این صورت، ماشین بردار پشتیبان، داده‌های اصلی را با به کارگیری مجموعه‌ای از توابع ریاضی که کرنل نام دارند در فضای جدیدی نگاشت و بازآرایی می‌کند که به این کار تبدیل (نگاشت) گفته می‌شود [۲۰].

کارآمدترین الگوریتم‌های داده‌کاوی به شمار می‌روند. این مدل‌ها علاوه بر دارا بودن یک شالوده ریاضیاتی منسجم در تئوری یادگیری آماری، عملکرد بسیار خوب و موفقیت آمیزی را در کاربردهای عملی و واقعی از خود نشان می‌دهند. یکی از ویژگی‌های قابل توجه الگوریتم مذکور آن است که در اغلب موارد هیچ حساسیتی را نسبت به ابعاد داده‌ها از خود نشان نمی‌دهند. در حالت کلی، ماشین‌های بردار پشتیبان بر پایه مفهوم صفحات تصمیم هستند که مرز تصمیم را تعریف می‌کنند، که یک صفحه تصمیم، داده‌های با برچسب‌های مختلف را از هم تفکیک می‌کند [۲۰]. در یک الگوریتم خطی‌سازی به کمک ماشین بردار



شکل ۴. شمایی از خطی‌سازی به کمک الگوریتم ماشین بردار پشتیبان با در نظر گرفتن انحراف ϵ

نماید. ایجاد چنین شرایطی با حل معادله بهینه‌سازی زیر قابل دستیابی است:

(۲)

$$\text{Minimize} = \frac{1}{2} \|w\|^2 \quad \text{subject to} \quad \begin{cases} y_i - (w, x_i) - b \leq \epsilon \\ (w, x_i) + b - y_i \leq \epsilon \end{cases}$$

که در آن w ، b ، X_i و Y_i همان پارامترهای معادله (۲) بوده و ϵ نیز میزان انحراف اعمال شده می‌باشد. نهایتاً برای تبدیل معادله بهینه‌سازی حاوی یک مجموعه قیود

اگر داده‌های آموزش به صورت k سری به صورت $(x_1, y_1), \dots, (x_k, y_k)$ باشند، تابع خطی‌سازی مربوطه به صورت رابطه (۲) قابل بیان است:

$$F(x) = (w, x) + b \quad \text{with } (w, x) \in R \quad (1)$$

که در آن x بردار ورودی، w وزن بردار و b میزان اختلال می‌باشد. تابع $f(x)$ همواره می‌بایست به گونه‌ای تعیین گردد که به طور هم‌زمان میزان کمترین انحراف ϵ تعیین شده و همچنین تابع، مقدار مناسب w را اختیار

$$k(x_i, x_j) = \frac{1}{\left[1 + \left(2\sqrt{\|x_i, x_j\|^2} \sqrt{2^{1/w} - 1/\sigma}\right)^2\right]^{1/w}} \quad (۸)$$

۲.۴. k-نزدیک‌ترین همسایگی

این تکنیک بر اساس مفهوم تشابه شکل گرفته است. نتایج حاصل از استدلال مبتنی بر حافظه^۵، بر پایه موفقیت های مشابهی که در گذشته اتفاق افتاده، بنا نهاده شده است. بدین معنی که این الگوریتم، تمامی موارد موجود را ذخیره کرده و بر اساس اندازه‌گیری شباهت‌ها به پیش‌بینی عددی هدف می‌پردازد [۱۹]. الگوریتم انجام پیش‌بینی با استفاده از روش k-نزدیک‌ترین همسایگی به صورت زیر است:

بردار سطری m ستونه مقادیر متغیرهای پیش‌بینی‌کننده x_j در زمان t به صورت رابطه (۱۰) می‌باشد:

$$Pr_{jt} = (x_{jt}) \quad j = 1 \dots m \quad (۹)$$

ماتریس m ستونه n سطری از مقادیر متغیرهای پیش‌بینی‌کننده x_j در سری زمانی تاریخی به صورت زیر است.

$$Pr_{j,(t-i)} = (x_{j,(t-i)}) \quad j = 1 \dots m, i = 1 \dots n \quad (۱۰)$$

با استفاده از تابع فاصله سنج، فواصل بین بردار Pr_{jt} با سطرهای ماتریس $Pr_{j,(t-i)}$ استخراج می‌گردد.

$$Dist(t-i) = f(w_j, x_{j,(t-i)}, x_{jt}) \quad (۱۱)$$

که در این رابطه شاخص j نشان‌دهنده متغیرهای پیش‌بینی‌کننده و شاخص i بیان‌کننده گام زمانی در سری تاریخی است. مقادیر w_j وزن‌هایی است که برای پیش‌بینی‌کننده‌ها، در نظر گرفته می‌شود [۱]. به طور کلی می‌توان برای بهبود عملکرد مدل‌های k-نزدیک‌ترین همسایگی اقداماتی از جمله انتخاب روشی جهت تخمین

نامعادلات و نامساوی به یک معادله صریح با لحاظ نمودن پارامترهای لاگرانژ (λ'_i, λ_i) معادله (۴) به دست می‌آید.

$$f(x, b) = \sum_{i=1}^k (\lambda'_i - \lambda_i)(x_i, x) + b \quad (۳)$$

حال معادله خطی‌سازی بالا، با به کارگیری توابع کرنل، برای مسائل رگرسیون غیر خطی نیز قابل استفاده خواهد شد [۲۴]. لذا در نهایت معادله تابع رگرسیون، به شکل رابطه (۵) بازنویسی می‌شود.

$$f(x, b) = \sum_{i=1}^k (\lambda'_i - \lambda_i)k(x_i, x) + b \quad (۴)$$

که در رابطه (۵)، $K(x_i, x)$ نمایانگر تابع کرنل می‌باشد. از انواع توابع کرنل می‌توان به توابع خطی^۱، چندجمله‌ای^۲، تابع شعاع محور^۳ و پیرسون^۴ اشاره کرد که این توابع در روابط (۶) تا (۹) ارائه شده‌اند. در این تحقیق برای مدل‌سازی میزان بار معلق رودخانه ليقوان چای توسط رگرسیون بردار پشتیبان، از نرم افزار Statistica استفاده شده است.

(۵) تابع کرنل چند جمله‌ای ساده

$$k(x_i, x_j) = (x_i^T x_j + 1)^p$$

(۶) تابع کرنل چند جمله‌ای نرمال شده

$$k(x_i, x_j) = \frac{(x_i^T x_j + 1)^p}{\sqrt{(x_i^T x_i)(x_j^T x_j)}}$$

(۷) تابع پایه شعاعی

$$k(x_i, x_j) = \exp\left(-\gamma|x_i - x_j|^2\right)$$

¹- Linear

²-Poly kernel

³-Radial Basis Function

⁴-Pearson

⁵- Memory-Based Reasoning

۵.۲. معیارهای ارزیابی مدل‌ها

عملکرد مدل‌های رگرسیون بردار پشتیبان و k -نزدیک‌ترین همسایگی در این تحقیق بر پایه محاسبه ضریب همبستگی^۵، نش-ساتکلیف^۶، جذر میانگین مربعات خطا^۷، و میانگین خطای مطلق^۸ مورد ارزیابی قرار گرفت. فرمول‌های محاسبه آماره‌های فوق در روابط (۱۶) تا (۱۹) ارائه گردیده است. در این روابط مقادیر X شامل مقادیر مشاهداتی و مقادیر Y شامل مقادیر محاسباتی می‌باشند.

$$RMSE = \sqrt{\frac{\sum_{i=0}^n (y_i - x_i)^2}{N}} \quad (15)$$

$$R_{pearson} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (16)$$

$$MAE = \frac{1}{n} \sum_{i=0}^n |X_i - Y_i| \quad (17)$$

$$E = 1 - \frac{\sum_{i=1}^n (X_i - Y_i)^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \quad (18)$$

۳. نتایج

بعد از انجام آزمون‌های همگنی نرمال استاندارد، ون نیومن، بیشند و پتیت بر روی داده‌های موجود، ملاحظه گردید داده‌ها از شاخص‌های همگنی بالایی برخوردار بوده و برای مدل‌سازی می‌توانند مورد استفاده قرار گیرند. از

بهترین همسایه‌ها، توسعه توابع انتقال اطلاعات و توسعه توابع فاصله‌سنجی انجام داد. برای تخمین بهترین همسایگی‌ها در روش k -نزدیک‌ترین همسایگی، روش‌های مختلفی ارائه شده‌اند که بسته به دقت مورد استفاده و پیچیدگی و حجم مسئله قابل استفاده هستند. از جمله این روش‌ها استفاده از رابطه تجربی $K = \sqrt{n}$ می‌باشد که در این رابطه n طول سری زمانی و k بهترین تعداد همسایگی مورد استفاده در این روش است. میزان کارایی این رابطه با افزایش طول سری زمانی افزایش می‌یابد. سعی و خطا روش دیگری است که می‌توان در این زمینه به کار برد بدین ترتیب که با انتخاب همسایگی‌های مختلف در محدوده بهترین همسایگی و استخراج خطاهای پیش‌بینی، سعی در پیدا کردن بهترین همسایگی‌ها را داشت. قدم بعدی انتخاب تابعی برای انتقال اطلاعات می‌باشد که وظیفه این تابع همگن کردن فضای اطلاعات ورودی مدل به جهت جلوگیری از خطاهای مربوطه است. تابع دیگری که نقش اساسی در عملکرد روش k -نزدیک‌ترین همسایگی دارا می‌باشد، تابع فاصله‌سنجی است که از جمله این توابع می‌توان به توابع زیر اشاره کرد [۱].

$$\sqrt{\sum_{i=1}^k (x_i - y_i)^2} \quad (12) \text{ اقلیدسی}^1$$

$$\left(\sum_{i=1}^k (|x_i - y_i|)^q \right)^{1/q} \quad (13) \text{ مینکوسکی}^2$$

$$\sqrt{\sum_{i=1}^k |x_i - y_i|} \quad (14) \text{ منهتن}^3$$

در این تحقیق جهت مدل‌سازی میزان بار معلق رودخانه ليقوان چای، توسط روش K -نزدیک‌ترین همسایگی از نرم‌افزار Weka که در دانشگاه ویکاتو^۴ نیوزلند توسعه یافته، استفاده شده است.

1- Euclidean

2- Minkowski

3- Manhattan

4- Waikato

5- Correlation coefficient

6- Nash-Sutcliffe coefficient (E)

7- Root Mean Square Error

8- Mean Absolute Error

شود. همچنین هنگام مدل‌سازی و برآورد میزان رسوب توسط روش k-نزدیک‌ترین همسایگی پس از انجام سعی و خطاهای متعدد، نتیجه بهینه زمانی حاصل شد که از تعداد یک همسایه نزدیک و تابع اقلیدسی به‌عنوان تابع فاصله‌سنجی بهینه استفاده گردید. نتایج مربوط به روش‌های رگرسیون بردار پشتیبان و k-نزدیک‌ترین همسایگی در جدول (۲) ارائه گردیده است.

میان کل مجموعه داده‌ها تنها تعداد دو داده مغشوش (پرت) وجود داشت که حذف گردیدند.

هنگام مدل‌سازی و برآورد میزان رسوب توسط رگرسیون بردار پشتیبان بعد از سعی و خطاهای متعدد و بررسی انواع توابع کرنل و هایپرپارامترهای مربوطه، مشخص گردید نتیجه بهینه زمانی حاصل می‌شود که تابع پیرسون به عنوان تابع کرنل مورد استفاده قرار گرفته و مقادیر $(C = 1/5)$ و $(\omega = 0/3)$ و (σ) در نظر گرفته

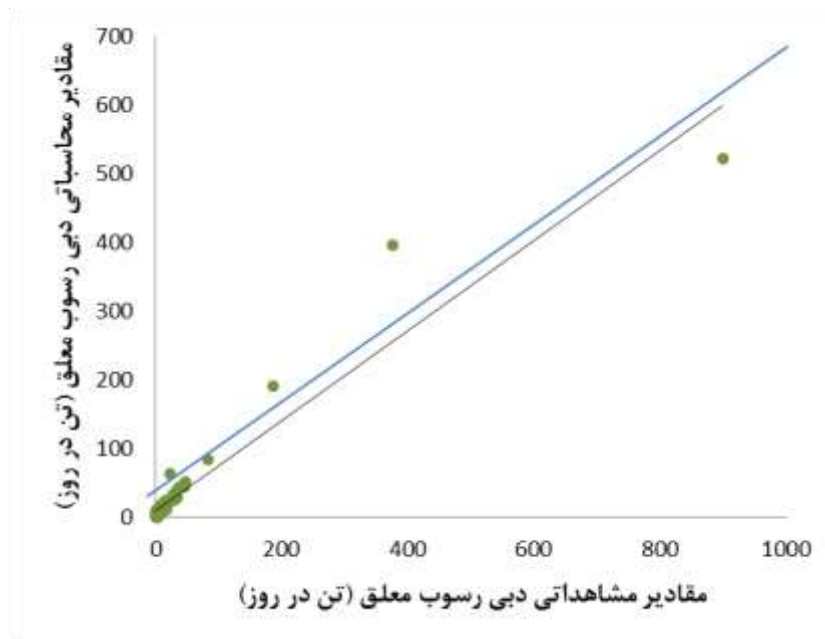
جدول ۲. نتایج آماری به دست آمده برای برآورد بار معلق رودخانه ليقوان چای

آماره‌های ارزیابی روش استفاده شده	ضریب همبستگی	ضریب نش	ریشه میانگین مربعات خطا (تن در روز)	میانگین خطای مطلق (تن در روز)
رگرسیون بردار پشتیبان	۰/۹۵۹	۰/۸۴۷	۴۳/۵۴۷	۹/۱۷۶
k-نزدیک‌ترین همسایگی	۰/۸۸۹	۰/۶۳۸	۶۶/۴۷۹	۱۱/۶۲۳

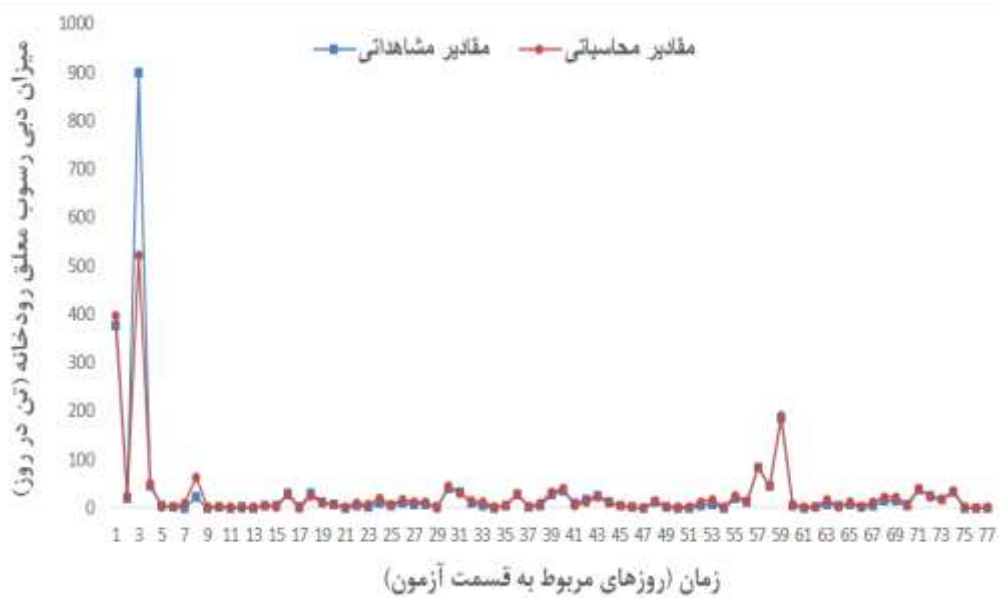
انطباق مناسبی بین داده‌های مشاهداتی و محاسباتی مشاهده می‌گردد. همچنان که در اشکال (۵) و (۶) مشاهده می‌شود بیشترین تطابق بین بهترین برازش بین نقاط و نیمساز ربع اول مربوط به داده‌های به دست آمده از روش رگرسیون بردار پشتیبان است که نشان‌دهنده دقت بالای این مدل در زمینه برآورد بار معلق رودخانه ليقوان چای و نزدیکی و همخوانی بیشتر مقادیر پیش‌بینی شده و مقادیر واقعی میزان دبی رسوبی معلق می‌باشد. روش رگرسیون بردار پشتیبان توانسته انطباق خوبی بین مقادیر محاسباتی و مشاهداتی میزان بار معلق رودخانه ليقوان چای در تمامی شرایط ایجاد نماید که نشان‌دهنده توانایی این روش در حل مسائل غیر خطی بوده، که این مسئله به علت استفاده این روش از تابع کرنل می‌باشد.

همچنان که در جدول (۲) مشاهده می‌شود مدل رگرسیون بردار پشتیبان با ارائه مقادیر بالاتر ضریب همبستگی و نش و مقادیر کمتر آماره‌های مربوط به خطا (ریشه میانگین مربعات خطا و میانگین خطای مطلق) دارای دقت بسیار بالا و قابل توجهی در زمینه برآورد میزان دبی رسوب معلق رودخانه ليقوان چای می‌باشد. برای درک بهتر نتایج مربوطه نمودار داده‌های محاسباتی به دست آمده از روش‌های مختلف نسبت به داده‌های مشاهداتی موجود، در شکل‌های (۵) و (۶) نشان داده شده است.

بررسی دقیق نتایج نشان‌دهنده این مطلب است که هر دو روش بررسی شده در این مطالعه دارای عملکرد نسبتاً مناسب و قابل قبولی در زمینه برآورد میزان دبی بار رسوبی معلق رودخانه می‌باشند و در هر دو روش

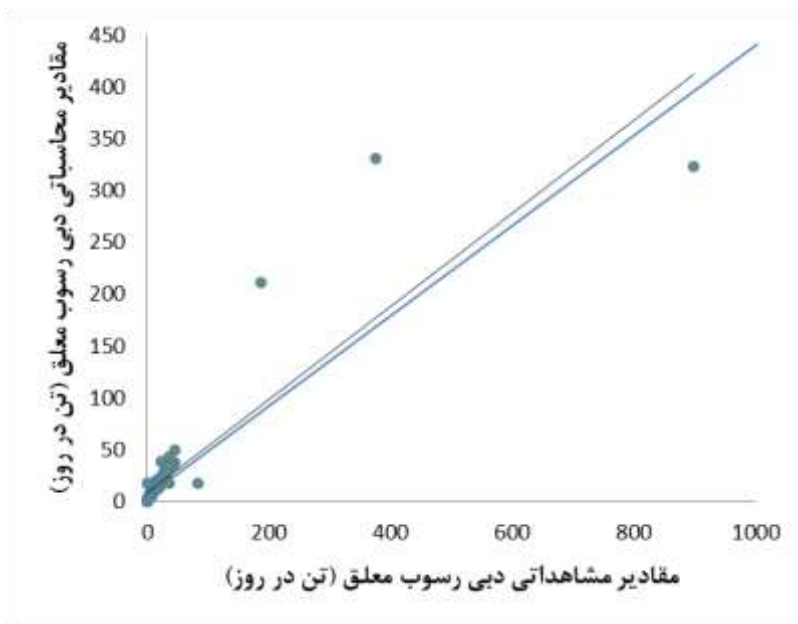


(الف)

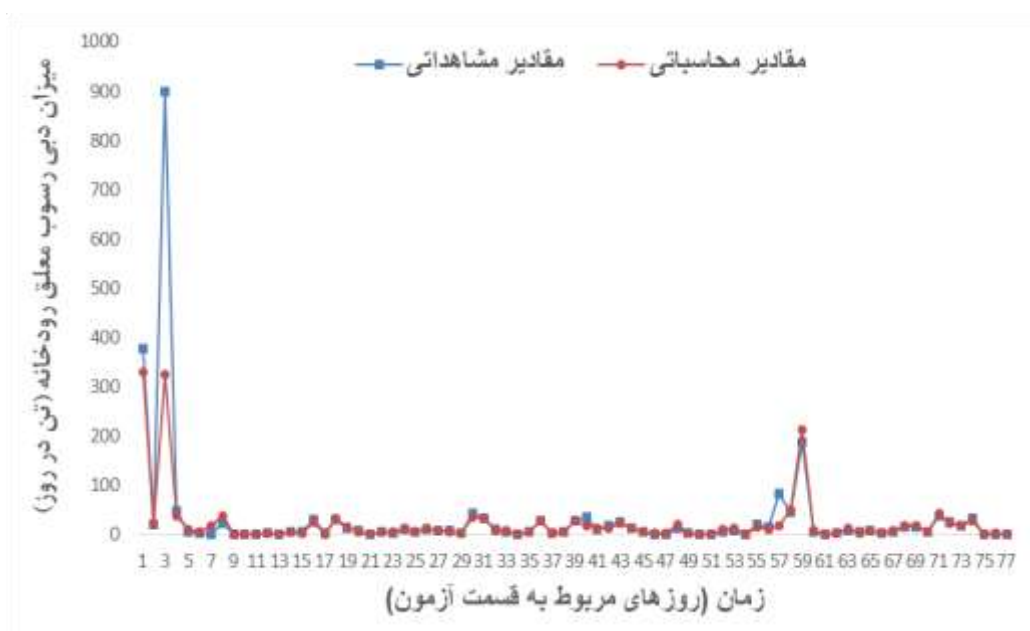


(ب)

شکل ۵. مقادیر محاسباتی و مشاهداتی میزان بار معلق به دست آمده از رگرسیون بردار پشتیبان (الف) نمودار پراکنش، (ب) نمودار سری زمانی



(الف)



(ب)

شکل ۶. مقادیر محاسباتی و مشاهده‌ای میزان بار معلق به دست آمده از k-نزدیک‌ترین همسایگی (الف) نمودار پراکنش، (ب) نمودار سری زمانی

۴. بحث و نتیجه گیری

طراحی بهینه و عملکرد مناسب سازه‌های منابع آب، نظیر مخزن، سد و کانال، نیازمند تخمین دقیق از بار رسوبی رودخانه است. در این تحقیق پس از جمع‌آوری داده‌های مورد نیاز مربوط به ليقوان چای، این داده‌ها توسط آزمون‌های همگنی شامل نرمال استاندارد، ون نیومن، بیشند و پتیت مورد بررسی قرار گرفتند و پس از پالایش داده‌ها نسبت به مدل‌سازی میزان دبی رسوبی معلق توسط روش‌های رگرسیون بردار پشتیبان، k-نزدیک ترین همسایگی اقدام گردید.

مقایسه نتایج به دست آمده از سایر مطالعات می‌تواند کمک شایانی برای ارزیابی عملکرد روش‌های مورد استفاده در این مطالعه بنماید. در تحقیق مشابهی [۶] بار معلق رسوب روزانه رودخانه دویرج با استفاده از دو مدل شبکه عصبی مصنوعی و ماشین بردار پشتیبان پیش‌بینی گردید. نتایج نشان داد شبکه عصبی مصنوعی با ارائه ضریب همبستگی برابر با ۰/۹۹ و مقدار ریشه میانگین مربعات خطا را برابر ۰/۳۴ (تن در روز) و ماشین بردار پشتیبان با ارائه ضریب همبستگی برابر با ۰/۹۸ و مقدار ریشه میانگین مربعات خطا را برابر ۰/۹۶ (تن در روز) عملکرد مناسبی در این زمینه داشته‌اند. دقت بالای نتایج این تحقیق و بخصوص مقادیر پایین آماره ریشه میانگین مربعات خطا می‌تواند مربوط به محدوده داده‌های استفاده شده در این تحقیق باشد. متأسفانه اطلاعاتی در مورد محدوده داده‌های استفاده شده در تحقیق مذکور موجود نمی‌باشد لذا امکان مقایسه دقیق و صحیحی بین نتایج این تحقیق با نتایج مطالعه حاضر وجود ندارد. در پژوهشی [۸] رسوبات معلق رودخانه قطورچای با استفاده از شبکه عصبی مصنوعی و منحنی سنجه رسوب پیش‌بینی گردید. نتایج نشان داد شبکه عصبی مصنوعی هنگام مدل‌سازی رسوبات معلق این رودخانه، ضریب همبستگی را برابر با ۰/۸۷۳ و مقدار ریشه میانگین مربعات خطا را برابر ۶۲۲ (تن در روز) ارائه کرده است.

همچنین در تحقیق دیگری [۲۳] بار رسوب جریان را در حوزه آبخیز سمندگان با استفاده از شبکه عصبی مصنوعی پیش‌بینی کردند که در بهترین حالت ضریب همبستگی برابر با ۰/۹۲۸ و ریشه میانگین مربعات خطا برابر با ۱/۸۱۵ (تن در روز) به دست آوردند. در مطالعه‌ای [۱۷] مقدار دبی بار معلق رودخانه اهرچای با استفاده از روش‌های مدل درختی M5 و رگرسیون بردار پشتیبان مدل‌سازی گردید. نتایج نشان داد مدل درختی M5 با ارائه ضریب همبستگی برابر با ۰/۹۲ و مقدار ریشه میانگین مربعات خطا را برابر ۱۱۵۲/۲۳ (تن در روز) و رگرسیون بردار پشتیبان با ارائه ضریب همبستگی برابر با ۰/۸۲ و مقدار ریشه میانگین مربعات خطا را برابر ۱۷۲۴/۷۴ (تن در روز) عملکرد قابل قبولی در این زمینه داشته‌اند. در حالیکه در تحقیق حاضر روش رگرسیون بردار پشتیبان با ارائه ضریب همبستگی ۰/۹۵۹ و ریشه میانگین مربعات خطا برابر ۴۳/۵۴۷ (تن در روز) و روش k-نزدیک‌ترین همسایگی با ارائه ضریب همبستگی ۰/۸۸۹ و ریشه میانگین مربعات خطا برابر ۶۶/۴۷۹ (تن در روز) توانسته‌اند به برآورد میزان دبی بار معلق در رودخانه ليقوان چای بپردازند. همچنان که قبلاً نیز اشاره شد اختلاف در میزان ریشه میانگین مربعات خطا در مطالعات مختلف می‌تواند مربوط به تفاوت در نوع و بازه‌ی داده‌های مورد استفاده در مطالعات مختلف باشد. مقایسه نتایج بدست آمده از این مطالعه با مطالعات مشابه نشان دهنده عملکرد مناسب و قابل قبول روش‌های به کار رفته در این مطالعه می‌باشد.

بررسی دقیق نتایج نشان داد علی‌رغم عملکرد نسبتاً مناسب هر دو روش داده‌کاوی استفاده شده در این مطالعه، روش رگرسیون بردار پشتیبان که به دلیل استفاده از توابع کرنل قابلیت حل مسائل غیر خطی را دارد، دقت بیشتری نسبت به روش k-نزدیک‌ترین همسایگی دارا است. لذا استفاده از این روش در موارد مشابه توصیه می‌گردد.

References

- [1]. Azmi, M. and Araghi nejed, Sh. (2011). Development of K-Nearest Neighbor Regression Method in Forecasting River Stream Flow, *Journal of Water & Wastewater*, 2, 108-119. (In Persian)
- [2]. Dehgani, A.A., Malek Mohammadi, M. and Hezarjaribi, A. (2010). Estimation of Suspended Sediment Load in Behesht Abad River by Using Artificial Neural Network, *Journal of Water and Soil Conservation*, 17(1), 159-168. (In Persian)
- [3]. Eder, A., Strauss a, P., Krueger b, T. and Quinton b, J.N. (2010). A Comparative calculation of suspended sediment loads with respect to hysteresis effects (in the Petzenkirchen catchment, Austria), *Journal of Hydrology*, 389, 168-176.
- [4]. Falamaki, A., Eskandari, M., Baghlani, A. and Ahmadi, S.A. (2013). Modeling total sediment load in rivers using artificial neural networks, *journal of water and soil conservation*, 2(3), 13-25. (In Persian)
- [5]. Iadanza, C. and Napolitano, F. (2006). Sediment transport time series in the Tiber River, *Physics and Chemistry of the Earth*, 31, 1212-1227.
- [6]. Kakaei Lafdani, E., Moghaddam Nia, A. and Ahmadi, A. (2013). Daily Suspended Sediment Load prediction Using Artificial Neural Networks and Support Vector Machines, *Hydrology*, 478, 50-62.
- [7]. Kao, Sh., Lee, T. and Milliman, J.D. (2005). Calculating highly fluctuated suspended sediment fluxes from mountainous rivers in Taiwan, *TAO*, 16(3), 653-675.
- [8]. Khazaie Poul, A. and Talebi, A. (2013). Investigation of Possibility of Suspended Sediment Prediction Using The Combination of Sediment Rating Curve and Artificial Neural Network (Case Study: Ghatorchai River, Yazdakan Bridge), *Quarterly Journal of Environmental Erosion Researches*, 2(9), 73-82. (In Persian)
- [9]. Kia, E., Emadi, A.R. and Fazlola, R. (2013). Investigation and Evaluation of Artificial Neural Networks in Babolroud River Suspended Load Estimation, *Journal of Civil Engineering and Urbanism*, 3(4), 183-190.
- [10]. Nourani, V. (2014). A Review on Applications of Artificial Intelligence-Based Models to Estimate Suspended Sediment Load, *International Journal of Soft Computing and Engineering (IJSCE)*, 3(6), 121-127.
- [11]. Onderka, M., Krein, A. and Wrede, S. (2012). Dynamics of storm-driven suspended sediments in a headwater catchment described by multivariable modeling, *Journal of Soils Sediments*, 12, 620-635.
- [12]. Ozturk, F., Apaydin, H. and Walling, D.E. (2001). Suspended Sediment loads through flood events for stream of sakarya Basin, *Turkish Journal of Engineering and Environmental Sciences*, 25, 643-650.
- [13]. Partal, T. and Cigizoglu, H.K. (2008). Estimation and forecasting of daily suspended sediment data using wavelet neural networks, *Journal of Hydrology*, 358, 317-331.
- [14]. Rajabi, M., Feizollahpour, M. and Roustae, S. (2015). Using NDE model for estimation of suspended sediment load in comparison with ANFIS and RBF case study: Givi Chay, *Geography and Development Iranian Journal*, 39(2), 1-16. (In Persian)
- [15]. Rastgar, H. and Habibi, M. (2011). Assessment of five sediment estimation methods in the Jegin River in Hormozgan province, *Journal of Engineering and watershed management*, 3(1), 12-20. (In Persian)
- [16]. Sattari, MT., Rezazadeh Joudi, A. and Kusiak, A. (2015). Estimation of water quality parameters with data-driven models, *American Water Works Association*, 108(4): 232-239.
- [17]. Sattari, MT., Rezazadeh Joudi, A., Safdari, F. and Ghahramanzadeh, F. (2016). Performance Evaluation of M5 Tree Model and Support Vector Regression Methods in Suspended Sediment Load Modeling, *Journal of Water and Soil Resources Conservation*, 6(1), 109-124. (In Persian)
- [18]. Shafaie, B. (2011). *Sediment hydraulic*, Shahid Chamran University, Press. (In Persian)
- [19]. Shahrabi, J. (2013). *Data mining 2*, Tehran, Industrial university of amirkabir, Jahad daneshgahi Press. (In Persian)

- [20]. Shahrabi, J., Hejazi, T.H. (2011). Data mining. Tehran, Industrial University of amirkabir, Jahad daneshgahi Press. (In Persian)
- [21]. Tabatabaei, M., Solaimani, K., Habibnejad Roshan, M. and Kavian, A. (2014). Estimation of Daily Suspended Sediment Concentration using Artificial Neural Networks and Data Clustering by Self-Organizing Map (Case Study: Sierra Hydrometry Station- Karaj Dam Watershed), Journal of Watershed Management Research, 5(10), 98-116. (In Persian)
- [22]. Vali, A., Moayeri, M., Ramsht, M.H. and Movahedinia, N. (2010). Analysis and Comparison of artificial neural networks and regression models in suspended sediment Prediction case study: Eskandari Catchment Area located in Zayanderood Basin, journal of Physical Geography Research Quarterly, 71(1), 21-30. (In Persian)
- [23]. Vali, A.A., Ramesht, A., Seif, A. and Ghazavi, R. (2010). An assessment of the Artificial Neural Networks technique to geomorphologic modeling sediment yield (Case study Samandegan river system), Geography and Environmental Planning Journal, 44(4), 19-34. (In Persian)
- [24]. Vapnik, V. N. (1995). The nature of statistical learning theory, Newyork: springer-verlag.
- [25]. Yang, C.T., Marsooli, R. and Aalami, M.T. (2009). Evaluation of total load sediment transport formulas using ANN, International journal of Sediment Research, 24, 274-286.
- [26]. Zhou, Y., Lu, X.X., Huang, Y. and Zhu, Y.M. (2007). Suspended sediment flux modeling with artificial neural network: An example of the Longchuanjiang River in the upper Yangtze catchment, China. Journal of Geomorphology, 84, 111-125.