

نشریه مرتع و آبخیزداری، مجله منابع طبیعی ایران  
دوره ۶۷، شماره ۲، تابستان ۱۳۹۳

تاریخ دریافت: ۱۳۹۱/۸/۲۸

تاریخ پذیرش: ۱۳۹۲/۶/۱۹

ص ۲۸۹-۳۰۳

# ارزیابی تأثیر پیش‌پردازش متغیرهای ورودی به مدل ماشین بردار پشتیبان به روش آزمون گاما به منظور پیش‌بینی حجم

## رسوب معلق

- ❖ الهام کاکائی لعدانی؛ دانشجوی دکتری علوم و مهندسی آبخیزداری، دانشکده منابع طبیعی، دانشگاه تربیت مدرس، نور، ایران
- ❖ علیرضا مقدم‌نیا؛ دانشیار گروه احیای مناطق خشک و کوهستانی، پردیس کشاورزی و منابع طبیعی کرج، دانشگاه تهران، کرج، ایران
- ❖ آزاده احمدی؛ استادیار دانشکده مهندسی عمران، دانشگاه صنعتی اصفهان، اصفهان، ایران
- ❖ حیدر ابراهیمی؛ دانشجوی دکتری علوم و مهندسی آبخیزداری، گروه آبخیزداری، دانشگاه کاشان، کاشان، ایران

### چکیده

هدف از این مطالعه بررسی تأثیر پیش‌پردازش متغیرهای ورودی به روش آزمون گاما بر عملکرد مدل ماشین بردار پشتیبان جهت پیش‌بینی حجم رسوبات معلق رودخانه دویرج، واقع در استان ایلام، برای دوره ۱۹۹۴-۲۰۰۵ است. دبی جریان و بارندگی ورودی مدل و دبی رسوب معلق خروجی مدل در نظر گرفته شد. همچنین، طول دوره آموزش مدل با استفاده از آزمون گاما (Gamma Test (GT)) مشخص شد. سپس، به منظور بررسی تأثیر پیش‌پردازش متغیرهای ورودی بر عملکرد مدل، رسوب معلق با استفاده از مدل SVM، که هیچ‌گونه پیش‌پردازی بر روی متغیرهای ورودی آن صورت نگرفته بود، پیش‌بینی شد و نتایج با یکدیگر مقایسه گردید. نتایج نشان داد عملکرد مدل GT-SVM در مرحله آزمون با حداقل RMSE برابر با ۰/۹۶ تن در روز و حداکثر ضریب  $R^2$  برابر با ۰/۹۸ بین مقادیر پیش‌بینی شده و واقعی بهتر از عملکرد مدل SVM است.

واژگان کلیدی: آزمون گاما، ارزیابی کارایی، رسوب معلق، رودخانه دویرج، ماشین بردار پشتیبان.

## مقدمه

می‌تواند بار رسوب را بدون ایجاد مقادیر منفی پیش‌بینی کند [۳]. همچنین از مدل ماشین بردار پشتیبان برای پیش‌بینی سیل در یک حوزه آبخیز استفاده شد و نتایج مدل توسعه‌داده‌شده با مدل‌های مبنایی، همچون مدل ساده، روند، رگرسیون چندمتغیره، و شبکه عصبی، مقایسه شد [۷]. یکی از مسائل مهم در زمینه مدل‌سازی با استفاده از تکنیک‌های هوش مصنوعی انتخاب بهترین ترکیب از متغیرهای ورودی و تعیین طول دوره آموزش مناسب مدل‌هاست [۶]. بدین منظور تا کنون از روش‌های مختلف، مانند روش تجزیه مؤلفه‌های اصلی و الگوریتم ژنتیک، استفاده شده است. آزمون گاما<sup>۴</sup> یکی از روش‌هایی است که اخیراً برای این منظور هیدرولوژیست‌ها بدان توجه کرده‌اند [۱۵]. استفانسون<sup>۵</sup> و همکاران، در سال ۱۹۹۷، برای نخستین بار آزمون گاما را گزارش کردند و بعدها بسیاری از دانشمندان از آن استفاده و درباره آن بحث کردند [۲۰]. پیش‌بینی طوفان گرد و غبار با استفاده از تکنیک‌های هوش مصنوعی صورت گرفت و از GT برای تعیین بهترین ترکیب ورودی مدل استفاده شد [۹]. قابلیت تکنیک‌های آزمون گاما و نظریه آنتروپی برای تعیین متغیرهای مؤثر بر تابش خورشیدی در حوزه آبخیز Brue در انگلستان بررسی شد. نتایج نشان داد با استفاده از آزمون گاما تعداد متغیرهای مورد نیاز برای مدل‌سازی به طور قابل ملاحظه‌ای کاهش می‌یابد [۱]. نقش پیش‌پردازش پارامترهای ورودی به روش‌های PCA<sup>۶</sup>، GT، و روش رگرسیون گام به گام بر عملکرد مدل ماشین بردار پشتیبان برای پیش‌بینی جریان بررسی شد. نتایج نشان‌دهنده نقش مؤثر پیش‌پردازش در تعیین متغیرهای ورودی است

پیش‌بینی بار رسوب حمل‌شده توسط رودخانه‌ها در مطالعات مهندسی رودخانه امری ضروری است [۱۳]، [۱۴]. از دیدگاه مدیریت منابع آب، انتقال رسوب توسط رودخانه یکی از بزرگ‌ترین مشکلات بهره‌برداری از منابع آب سطحی در جهان مطرح است [۵]. به منظور پیش‌بینی بار رسوبات معلق، مدل‌های مختلفی توسعه یافته‌اند. این مدل‌ها، بر پایه روش‌های تجربی و آماری، در کنار مدل‌های عددی پیشنهاد شده‌اند و عملکرد آن‌ها بر پایه فرایندهای پیچیده فیزیکی تأثیرگذار در فرسایش و تولید رسوب معلق است [۱۸]. در دهه‌های اخیر، توسعه تکنیک‌های هوش مصنوعی، مانند شبکه‌های عصبی مصنوعی، ماشین‌های بردار پشتیبان<sup>۱</sup> و نظایر آن، به عنوان پیش‌بینی‌کننده‌های پدیده‌های هیدرولوژیکی، تحول عظیمی در پیش‌بینی‌ها به وجود آورده است [۱۱]، [۲۳]. از دیدگاه ریاضی، مدل ماشین بردار پشتیبان در گستره الگوریتم‌های دسته‌بندی و رگرسیون قرار می‌گیرد که با استفاده از اصول تئوری یادگیری آماری واپنیک<sup>۲</sup> فرموله می‌شود [۲۲]. این مدل به طور موفقیت‌آمیزی در زمینه‌های گوناگون، از جمله منابع آب و پیش‌بینی بار رسوب، به کار رفته است [۱۲]. در مطالعه‌ای از ماشین بردار پشتیبان برای مدل‌سازی رواناب استفاده شد [۲]. برای پیش‌بینی بار رسوب معلق رودخانه کائوپینگ<sup>۳</sup> تایوان از ماشین بردار پشتیبان و شبکه عصبی استفاده شد [۱۰]. نتایج نشان داد مدل ماشین بردار پشتیبان از توانایی بیشتری نسبت به مدل شبکه عصبی برخوردار است. با استفاده از مدل SVM و کرنل گوسی غلظت رسوب معلق پیش‌بینی شد. نتایج این مطالعه نشان داد مدل SVM

4. Gamma Test (GT)  
5. Stefansson  
6. Principal Component Analysis (PCA)

1. Support Vector Machine (SVM)  
2. Vapnik  
3. Kaoping

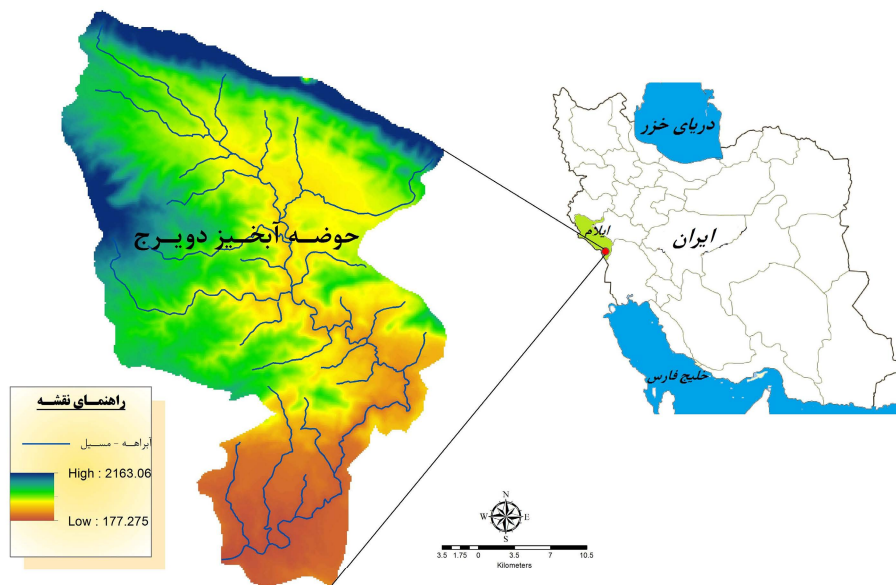
دهلران و آبدانان واقع شده است (شکل ۱). محدوده مورد مطالعه دارای مختصات جغرافیایی  $۱۶^{\circ}$  و  $۴۷^{\circ}$  تا  $۴۰^{\circ}$  و  $۴۷^{\circ}$  طول شرقی و  $۳۴^{\circ}$  و  $۳۲^{\circ}$  تا  $۰۵^{\circ}$  و  $۳۳^{\circ}$  عرض شمالی است که از شمال به کبیرکوه، از جنوب به موسیان، از شرق به کوه دال پری و حوزه آبخیز رودخانه چیخواب، و از غرب به دینار کوه و دهلران محدود می‌شود. متوسط بارندگی سالانه در این حوزه آبخیز در حدود  $۴۳۴/۴۳$  میلی‌متر است. میانگین حداکثر درجه حرارت در این حوزه آبخیز در حدود  $۲۷/۶۹$  درجه سانتی‌گراد و میانگین حداقل  $۱۳/۷۲$  درجه سانتی‌گراد است. جدول ۱ مشخصات ایستگاه‌های مورد استفاده را نشان می‌دهد.

[۱۷]. در این مطالعه، از مدل ماشین بردار پشتیبان رگرسیونی از نوع SVR-V به منظور پیش‌بینی حجم رسوبات معلق حوزه آبخیز رودخانه دویرج استفاده شد. در این راستا، از روش نوین آزمون گاما برای تعیین بهترین ترکیب ورودی مدل و طول دوره آموزش مدل استفاده شده است.

## ۲. روش شناسی

### ۱.۲ منطقه مورد مطالعه

منطقه مورد مطالعه به نام حوزه آبخیز رودخانه دویرج از رودخانه‌های مرزی کشور به‌شمار می‌رود. این حوزه آبخیز در محدوده استان ایلام و در شهرستان‌های



شکل ۱. موقعیت حوزه آبخیز دویرج در ایران و استان ایلام

جدول ۱. مشخصات ایستگاه‌های استفاده‌شده حوزه آبخیز دویرج

مشخصات جغرافیایی				نوع ایستگاه	نام ایستگاه
سال تأسیس	ارتفاع (متر)	عرض جغرافیایی	طول جغرافیایی		
۱۹۸۸	۹۴۰	$۳۲^{\circ} ۵۹'$	$۴۵^{\circ} ۲۵'$	باران‌سنجی	آبدانان
۱۹۷۲	۱۶۵	$۳۲^{\circ} ۳۵' ۴۰''$	$۴۷^{\circ} ۲۴'$	هیدرومتری	پل دویرج

بررسی همه این ترکیبات به زمان زیادی نیاز دارد. با استفاده از آزمون گاما می‌توان مؤثرترین متغیر در مدل‌سازی و بهترین ترکیب از متغیرهای ورودی را مشخص کرد. همچنین، با استفاده از  $M$  تست می‌توان طول دوره آموزش مدل پیش‌بینی برای ایجاد یک مدل هموار را مشخص نمود [۶].

### ۳.۲. ماشین بردار پشتیبان

ماشین‌های بردار پشتیبان دسته‌ای از روش‌های یادگیری با ناظر هستند که برای مسائل طبقه‌بندی و رگرسیونی به کار می‌روند. واپنیک و همکارانش، در سال ۱۹۹۵، این روش را معرفی کردند. این روش بر پایه نظریه یادگیری محاسباتی توسعه یافته است [۲۱، ۲۲]. در حالتی که داده‌ها به صورت خطی به شکل رابطه ۶ قابل ارائه‌اند:

$$y = \text{sign} \left( \sum_{i=1}^N y_i \alpha_i (X \bullet X_i) + b \right) \quad \text{رابطه ۶}$$

که در آن  $y$  خروجی معادله،  $y_i$  ارزش طبقه نمونه آموزشی  $X_i$ ، و  $\bullet$  نشان‌دهنده ضرب داخلی است. بردار  $x = (x_1, x_2, \dots, x_n)$  نشان‌دهنده یک داده ورودی و بردارهای  $X_i$ ،  $i = 1, \dots, N$  بردارهای پشتیبان‌اند. در رابطه ۶، پارامترهای  $\alpha_i, b$  تعیین‌کننده ابر صفحه هستند. اگر داده‌ها به صورت خطی تفکیک‌پذیر نباشند، رابطه ۶ به رابطه ۷ تغییر می‌یابد:

$$y = \text{sign} \left( \sum_{i=1}^N y_i \alpha_i K(X, X_i) + b \right) \quad \text{رابطه ۷}$$

تابع  $K(X, X_i)$  تابع کرنلی است که، برای ایجاد ماشین‌هایی با انواعی مختلف از سطوح تصمیم‌گیری غیرخطی در فضای داده‌ها، ضرب‌های داخلی تولید می‌کند. جدول ۲ چهار تبدیل استاندارد تابع کرنل را، که اغلب در مدل‌سازی و رگرسیون به کار می‌رود، نشان می‌دهد. انتخاب مناسب تابع کرنل امکان استفاده از یک تابع غیرخطی در فضای ورودی‌ها را برای تبدیل به تابع خطی در فضای ویژگی‌ها فراهم می‌کند.

### ۲.۲. آزمون گاما

آزمون گاما حداقل متوسط مجذور خطایی را که در مدل‌های غیرخطی پیوسته با داده‌های مشاهده‌نشده قابل حصول است تخمین می‌زند [۱۵]. فرض کنید یک سری داده مشاهداتی به صورت رابطه ۱ وجود دارد:

$$\text{رابطه ۱} \quad ((x_1, \dots, x_m), y) = (X, y)$$

که در آن  $X = (x_1, \dots, x_m)$  بردار ورودی در محدوده  $C \in \mathbb{R}^m$  است و  $y$  بردار خروجی. اگر رابطه ۲ بین اعضای جامعه برقرار باشد:

$$\text{رابطه ۲} \quad y = f(x_1, \dots, x_m) + \Gamma$$

$\Gamma$  در آن متغیر تصادفی است. آزمون گاما تخمینی برای واریانس خروجی مدلی غیرهموار است. آزمون گاما بر اساس  $N[i, k]$  است که شامل لیستی از  $k$  ( $1 \leq k \leq p$ ) امین همسایگی برای هر بردار  $X_i$  ( $1 \leq i \leq M$ ) است. تابع دلتا میانگین مجذور فاصله  $k$  امین همسایگی را محاسبه می‌کند [۱]:

$$\text{رابطه ۳} \quad \delta_M(k) = \frac{1}{M} \sum_{i=1}^M |X_{N[i, k]} - X_i|^2$$

که در آن  $| \cdot |$  نشان‌دهنده فاصله اقلیدسی<sup>۱</sup> است، تابع گاما متناظر آن:

$$\text{رابطه ۴} \quad \gamma_M(k) = \frac{1}{2M} \sum_{i=1}^M (y_{N[i, k]} - y_i)^2$$

که در آن مقدار  $y_{N[i, k]}$  مقدار  $y$  متناظر با  $k$  امین همسایگی  $X_i$  در رابطه ۴ است. به منظور محاسبه  $\Gamma$ ، یک خط رگرسیون خطی از  $p$  نقطه بر مقادیر  $\delta_M(k)$  و  $\gamma_M(k)$  برازش داده می‌شود.

$$\text{رابطه ۵} \quad \gamma = A\delta + \Gamma$$

عرض از مبدأ این خط  $\delta = 0$ ، مقدار  $\Gamma$  را نشان می‌دهد و  $\gamma_M(k)$  برابر با واریانس خطاهاست. اگر  $m$  تعداد متغیرهای ورودی باشد، در این صورت  $1 - 2^m$  ترکیب از میان آن‌ها وجود خواهد داشت.

1. Euclidean distance

جدول ۲. فرمول توابع مختلف کرنل

تابع کرنل	فرمول
Linear	$K(x, z) = x.z$
Polynomial	$K(x, z) = (1 + (x.z))^\alpha$
(RBF) Radial Basis Function	$K(x, z) = \exp(-\alpha x - z ^r)$
Exponential RBF	$K(x, z) = \exp(-\alpha x - z )$

## رابطه ۸

$$R^r = \frac{\sum_{i=1}^n (Q_{(s,\text{predicted})i} - \bar{Q}_{(s,\text{predicted})})(Q_{(s,\text{observed})i} - \bar{Q}_{(s,\text{observed})})}{\sqrt{\sum_{i=1}^n (Q_{(s,\text{predicted})i} - \bar{Q}_{(s,\text{predicted})})^2 (Q_{(s,\text{observed})i} - \bar{Q}_{(s,\text{observed})})^2}}$$

## رابطه ۹

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Q_{(s,\text{predicted})i} - Q_{(s,\text{observed})i})^2}$$

در این روابط  $Q_{(s,\text{observed})i}$  و  $Q_{(s,\text{predicted})i}$  به ترتیب مقادیر مشاهده‌شده و پیش‌بینی‌شده در زمان  $i$ ،  $\bar{Q}_{(s,\text{observed})}$  و  $\bar{Q}_{(s,\text{predicted})}$  میانگین مقادیر مشاهده‌شده و شبیه‌سازی‌شده، و  $n$  تعداد نمونه‌هاست. همچنین، به منظور بررسی تأثیر پیش‌پردازش متغیرهای ورودی بر عملکرد مدل SVM، خطای نسبی میان مقادیر ماکزیمم رسوب واقعی و رسوب معلق پیش‌بینی‌شده متناظر با آن بر اساس رابطه ۱۰ محاسبه شده است:

## رابطه ۱۰

$$E_{\text{Relative}} = \frac{Q_{(m,\text{predicted})i} - Q_{(m,\text{observed})i}}{Q_{(m,\text{observed})i}}$$

که در این رابطه  $Q_{(m,\text{observed})i}$  و  $Q_{(m,\text{predicted})i}$  به ترتیب ماکزیمم مقادیر رسوب معلق واقعی و پیش‌بینی‌شده در زمان  $i$  است.

## ۳. نتایج

برای پیش‌بینی رسوب معلق، ورودی‌های بارش و رواناب روزانه متناظر با رسوب و همچنین تأثیر مقادیر

در یک مدل رگرسیونی SVM لازم است وابستگی تابعی متغیر وابسته  $y$  به مجموعه‌ای از متغیرهای مستقل  $x$  تخمین زده شود. دو نمونه از مدل‌های SVM تعریف شده است که عبارت‌اند از: ۱. مدل‌های رگرسیونی SVM نوع اول که به مدل‌های  $v$ -SVM مشهورند؛ ۲. مدل‌های رگرسیونی که به  $\varepsilon$ -SVM مشهورند [۱۷]. مدل‌های  $v$ -SVM نسبت به مدل‌های  $\varepsilon$ -SVM دارای تفسیر بامعناتری هستند، زیرا در مدل‌های  $v$ -SVM،  $v$  نشان‌دهنده کران بالای خطا در کسری از نمونه‌های آموزشی و کران پایین بردارهای پشتیبان است. در آموزش مدل‌های  $\varepsilon$ -SVM، کیفیت برازش بر اساس فاصله  $\varepsilon$  از داده‌های واقعی مشخص می‌شود. در واقع، مقادیر کوچک‌تر  $\varepsilon$  موجب برازش نزدیک‌تر میان داده‌های آموزش می‌شود. در اکثر مطالعات کرنل، RBF بهترین کرنل برای شبیه‌سازی معرفی شده است [۴، ۱۶]. بنابراین، در این مطالعه از مدل  $v$ -SVM با تابع کرنل RBF استفاده شده است.

## ۴.۲. معیارهای ارزیابی عملکرد مدل SVM

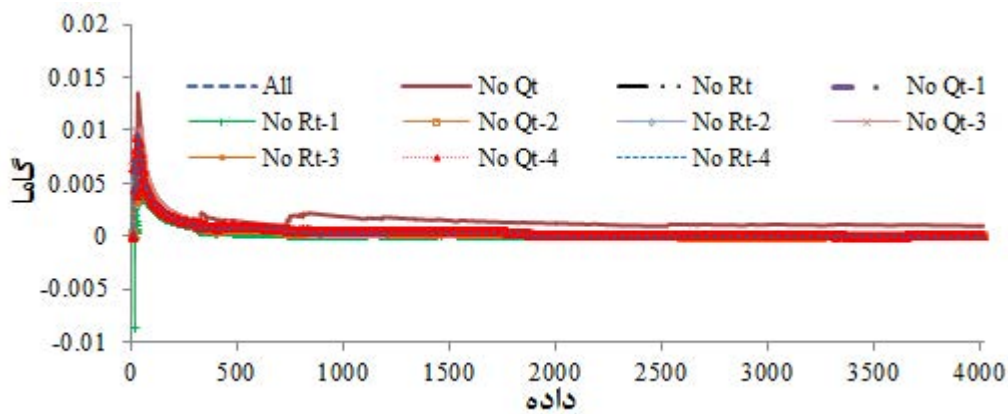
عملکرد مدل‌های پیش‌بینی با استفاده از معیارهای آماری ضریب همبستگی ( $R^2$ ) و ریشه میانگین مربعات خطا (RMSE) بررسی شده است. معیارهای آماری ذکر شده با استفاده از روابط ۸ و ۹ محاسبه می‌شوند:

موجود،  $Q_t$  بیشترین اثر را بر رسوب معلق دارد، زیرا حذف این پارامتر از مدل‌سازی باعث افزایش میزان گاما شده است. شکل ۲ نیز نشان می‌دهد که حذف متغیر  $Q_t$  از ترکیب ورودی به افزایش مقدار گاما منجر شده است، در نتیجه، منحنی مجانب متناظر با آن در فاصله بیشتری نسبت به سایر منحنی‌ها قرار گرفته است. حذف پارامترهای  $Q_{t-2}$ ،  $Q_{t-3}$  و  $R_{t-3}$  تأثیر یکسانی در افزایش میزان گاما داشته‌اند. علاوه بر این، تأثیر  $R_{t-2}$  و  $R_{t-3}$  کاملاً یکسان است، بنابراین، مؤثرترین متغیری که بایستی در ترکیب ورودی باشد متغیر  $Q_t$  است.  $R_t$ ،  $Q_{t-1}$  و  $R_{t-1}$  در جایگاه دوم اهمیت قرار دارند. متغیر  $Q_{t-2}$  بدون تأثیر بر روی میزان گاما است و از ترکیب ورودی حذف می‌شود. وجود یا فقدان سایر متغیرها در ترکیب ورودی به همراه متغیرهای مؤثر شناخته شده بایستی با استفاده از آزمون گاما بررسی شود. ترکیبات مختلف از وجود یا فقدان هر یک از ۴ متغیر مؤثر شناخته شده (و یا متغیرهایی که بایستی بررسی شوند، از قبیل  $Q_{t-2}$ ،  $R_{t-2}$ ،  $R_{t-3}$ ،  $Q_{t-3}$ ،  $R_{t-3}$ ) در ترکیب ورودی مدل پیش‌بینی با استفاده از آزمون گاما بررسی شد و برخی از ترکیبات، که دارای کمترین میزان گاما هستند، در جدول ۴ نشان داده شده است.

بارش و دبی جریان روزهای گذشته (تا ۴ روز قبل) بر مقدار رسوب معلق ارزیابی شد. به عبارت دیگر،  $Q_t, R_t, Q_{t-1}, R_{t-1}, Q_{t-2}, R_{t-2}, Q_{t-3}, R_{t-3}, Q_{t-4}, R_{t-4}$  ورودی مدل و  $Q_{st}$  خروجی در نظر گرفته شدند. در این بخش سعی بر آن است که پارامترهایی که وجود آن‌ها فقط بر پیچیدگی مدل می‌افزاید و تأثیر درخور توجهی در نتایج مدل ندارند شناسایی و حذف شوند. با استفاده از آزمون گاما، می‌توان متغیرهایی را با بیشترین تأثیرگذاری و کمترین تأثیرگذاری بر پیش‌بینی تعیین کرد. در این آزمون، نخست کل ترکیبات آزمون می‌شوند، سپس، به ترتیب یکی از متغیرها حذف می‌شود و این کار تا جایی صورت می‌گیرد که هر یک از متغیرها یک مرتبه کنار گذاشته شوند. هنگامی که در این ترکیبات مقدار گاما دارای بیشترین مقدار باشد متغیر حذف شده مهم‌ترین متغیر است و وقتی که این مقدار حداقل باشد متغیر کنار گذاشته شده کم‌اهمیت‌ترین متغیر است. جدول ۳ عدد گاما را به ازای حذف هر یک از متغیرها از ترکیب ورودی نشان می‌دهد. همچنین، شکل ۲ منحنی‌های مجانب مقادیر گاما متناظر با ۱۱ سناریوی مختلف تعریف شده در جدول ۳ را نشان می‌دهد. جدول ۳ نشان می‌دهد، از بین ۱۰ پارامتر

جدول ۳. تعیین مؤثرترین متغیر در پیش‌بینی بر مبنای آزمون گاما

سناریو	ترکیب ورودی	الگوی کار	گاما	سناریو	ترکیب ورودی	الگوی کار	گاما
۱	ALL	۱۱۱۱۱۱۱۱	-۰٫۰۰۰۰۴۱	۷	No $R_{t-2}$	۱۱۱۱۱۰۱۱۱۱	-۰٫۰۰۰۰۵۰
۲	No $Q_t$	۰۱۱۱۱۱۱۱۱	۰٫۰۰۰۰۹۵۷۵	۸	No $Q_{t-2}$	۱۱۱۱۱۱۰۱۱۱	-۰٫۰۰۰۰۴۲
۳	No $R_t$	۱۰۱۱۱۱۱۱۱۱	۰٫۰۰۰۰۲۵۲	۹	No $R_{t-3}$	۱۱۱۱۱۱۱۰۱۱	-۰٫۰۰۰۰۵۰
۴	No $Q_{t-1}$	۱۱۰۱۱۱۱۱۱۱	-۰٫۰۰۰۰۲۷	۱۰	No $Q_{t-4}$	۱۱۱۱۱۱۱۱۰۱	-۰٫۰۰۰۰۵۲
۵	No $R_{t-1}$	۱۱۱۰۱۱۱۱۱۱	-۰٫۰۰۰۰۱۱	۱۱	No $R_{t-4}$	۱۱۱۱۱۱۱۱۱۰	-۰٫۰۰۰۰۶۲
۶	No $Q_{t-2}$	۱۱۱۱۰۱۱۱۱۱	-۰٫۰۰۰۰۵۴				



شکل ۲. تغییرات مقادیر گاما متناظر با حذف هر یک از متغیرها از ترکیب ورودی

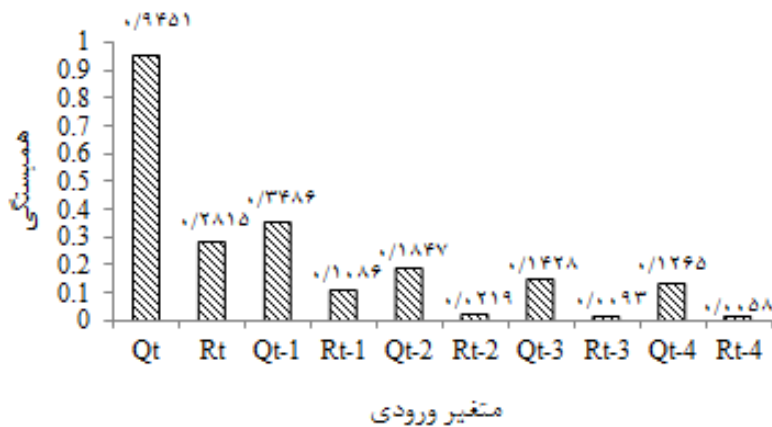
جدول ۴. تعیین بهترین ترکیب از متغیرهای ورودی به روش آزمون گاما

سناریو	ترکیب ورودی	الگوی کار	گاما ( $\Gamma$ )	خطای استاندارد (SE)
۱	$Q_t, Q_{t-2}, P_{t-2}$	۱۰۰۰۱۱۰۰۰۰	۰٫۰۰۰۰۳۰۶۲	۰٫۰۰۰۰۱۱۳۳
۲	$Q_t, P_t, P_{t-2}$	۱۱۰۰۰۱۰۰۰۰	-۰٫۰۰۰۰۲۱۰۵	۰٫۰۰۰۰۲۰۲۹۷
۳	$Q_t, P_t, Q_{t-2}$	۱۱۰۰۱۰۰۰۰۰	-۰٫۰۰۰۰۰۸۰۳	۰٫۰۰۰۰۱۰۵۶
۴	$Q_t, P_t, Q_{t-2}, P_{t-2}$	۱۱۰۰۱۱۰۰۰۰	$۵٫۷۲۳ \times 10^{-8}$	۰٫۰۰۰۰۲۰۸۳
۵	$Q_t, P_t$	۱۱۰۰۰۰۰۰۰۰	-۰٫۰۰۰۰۱۵۶۷۳	۰٫۰۰۰۰۷۴۰۲
۶	$Q_t, P_t, Q_{t-1}, P_{t-1}, Q_{t-2}, P_{t-2}$	۱۱۱۱۱۱۰۰۰۰	-۰٫۰۰۰۰۵۶۲۲	۰٫۰۰۰۰۳۲۱۳
۷	$Q_t, P_t, Q_{t-1}, Q_{t-2}, P_{t-2}$	۱۱۱۰۱۱۰۰۰۰	۰٫۰۰۰۰۰۱۰۱۴	۰٫۰۰۰۰۳۴۸۰
۸	$Q_t, P_t, P_{t-1}, Q_{t-2}, P_{t-2}$	۱۱۰۱۱۱۰۰۰۰	-۰٫۰۰۰۰۳۴۳۱	۰٫۰۰۰۰۳۰۹۶۲

رگرسیون بین متغیرهای ورودی و خروجی استفاده شده است.

شکل ۳ مقادیر همبستگی هر یک از متغیرهای ورودی با رسوب خروجی را نشان می‌دهد. همان‌طور که مشاهده می‌کنید،  $Q_t$ ،  $Q_{t-1}$  و  $R_t$  بیشترین همبستگی را با رسوب خروجی دارند. با توجه به متغیرهای شناخته‌شده توسط این روش، ۴ سناریوی مختلف از ترکیبات ورودی (جدول ۵) برای پیش‌بینی رسوب معلق تدوین شده است.

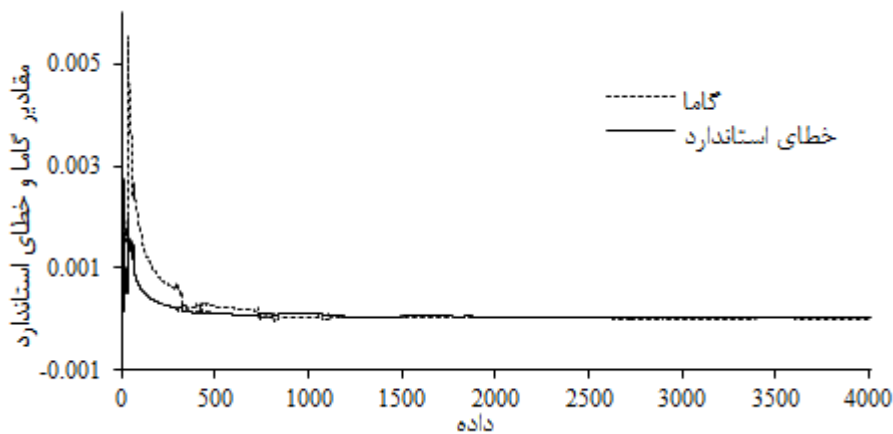
نتایج نشان می‌دهد، در میان ترکیبات بررسی‌شده، بهترین ترکیب ورودی از متغیرها زمانی است که فقط از بارش و دبی روز مورد نظر به همراه بارش و دبی با دو تأخیر زمانی (سناریوی ۴) استفاده شود ( $Q_t, R_t, Q_{t-2}, R_{t-2}$ ). مقدار کوچک گاما نشان می‌دهد که داده‌ها با ترکیب ارائه‌شده امکان رسیدن به نتایج بهتری در مدل‌سازی را میسر می‌سازد. همچنین، به منظور بررسی صحت نتایج روش آزمون گاما در تعیین بهترین ترکیب ورودی، از روش سستی



شکل ۳. مقادیر همبستگی میان ۱۰ متغیر ورودی و رسوب معلق

جدول ۵. ترکیبات تعیین شده از متغیرهای ورودی به روش همبستگی

سناریو	ترکیب ورودی	الگوی کار
۱	$Q_t, R_t, Q_{t-1}, R_{t-1}, Q_{t-2}, Q_{t-3}, Q_{t-4}$	۱۱۱۱۱۰۱۰۱۰
۲	$Q_t, R_t, Q_{t-1}$	۱۱۱۰۰۰۰۰۰۰
۳	$Q_t, Q_{t-1}, Q_{t-2}, Q_{t-3}, Q_{t-4}$	۱۰۱۰۱۰۱۰۱۰
۴	$Q_t, R_t, Q_{t-1}, Q_{t-2}$	۱۱۱۰۱۰۰۰۰۰



شکل ۴. آزمون M برای تعیین تعداد داده‌های آموزشی

آزمون زمانی که نمودار گاما به صورت نسبتاً صاف درآید، حداقل مقدار گاما، که دارای مقدار کمتری اشتباه استاندارد نیز باشد، نقطه مورد نظر مرز تقسیم‌بندی داده‌ها برای آموزش و آزمون شبکه

در این مطالعه، با استفاده از آزمون M، تعداد داده‌های مورد نیاز برای آموزش مدل مشخص گردیده است. این آزمون در ترکیبی که قرار است برای مدل‌سازی به کار رود صورت می‌گیرد. در این



مدل‌های SVM، GT-SVM، و Reg-SVM را نشان می‌دهد. با توجه به نتایج ارائه‌شده در جدول ۶، مشاهده می‌شود که تعیین پیش‌بینی‌کننده‌ها به روش رگرسیون بین متغیرهای ورودی و خروجی در ترکیب مدل ۳ (۱۰۱۰۱۰۱۰۱۰) و همچنین بدون پیش‌پردازش متغیرهای ورودی بهترین مدل در مرحله آموزش شبکه بوده است، زیرا دارای حداقل خطای ریشه متوسط مربعات (۰٫۶ تن در روز) بین مقادیر پیش‌بینی‌شده و واقعی است. ولی در مرحله آزمون مدل، با متغیرهای ورودی تعیین‌شده توسط آزمون گاما مدل GT-SVM دارای بهترین عملکرد (حداقل RMSE برابر با ۰٫۹۶ تن در روز) نسبت به سایر ترکیبات تعیین‌شده بوده است و این مطلب نشان‌دهنده کارایی مدل آزمون گاما در انتخاب متغیرهای ورودی برای مدل پیش‌بینی است. همچنین، پیش‌پردازش متغیرهای ورودی برای تعیین ترکیب ورودی باعث افزایش عملکرد مدل SVM شده است. به عبارت دیگر، پیش‌پردازش متغیرهای ورودی به روش آزمون گاما مقدار خطای پیش‌بینی در مرحله آزمون مدل را از مقدار ۲٫۸۸ تن در روز به مقدار ۰٫۹۶ تن در روز کاهش داده است و بهبود عملکرد مدل SVM را به همراه داشته است.

خواهد بود (شکل ۴). همان‌گونه که در شکل ۴ مشاهده می‌کنید، در نقطه ۲۹۵۰ نمودار تقریباً افقی می‌شود و این نقطه مرز تعداد داده‌های آموزش و آزمون مدل در نظر گرفته می‌شود. مقدار گاما و خطای استاندارد در این نقطه به ترتیب برابر با  $2,82 \times 10^{-6}$  و  $1,66 \times 10^{-5}$  است.

به منظور پیش‌بینی رسوب معلق، از مجموع ۴۰۱۴ داده، از ۲۹۵۰ داده برای آموزش مدل و از بقیه برای آزمون مدل استفاده شده است. برای پیش‌بینی با استفاده از مدل  $\nu$ -SVM و کرنل RBF باید مقادیر بهینه سه پارامتر  $C$ ،  $\varepsilon$ ، و  $\gamma$  تعیین شود. پارامترهای  $C$  و  $\varepsilon$  بر کیفیت و زمان آموزش تأثیر می‌گذارند. همچنین، مقدار پارامتر  $\gamma$  می‌تواند در وقوع بیش‌برازش و دون‌برازش شبکه تأثیرگذار باشد. در این مطالعه، مقادیر این پارامترها با استفاده از روش سعی و خطا تعیین شده است. در انتخاب مقادیر بهینه پارامترها سعی شده است که مدل با استفاده از این مقادیر دارای کمترین خطا در مرحله آزمون باشد. پارامترهای  $C$ ،  $\varepsilon$ ، و  $\gamma$  در مدل GT-SVM به ترتیب برابر ۹، ۰٫۰۰۹، و ۱ و در مدل SVM به ترتیب برابر ۳، ۰٫۰۰۷، و ۱٫۵ انتخاب شدند.

جدول ۶ نتایج حاصل از پیش‌بینی با استفاده از

جدول ۶. عملکرد مدل SVM بر اساس ترکیبات مختلف تعیین‌شده توسط روش همبستگی و آزمون گاما

آزمون		آموزش										
Reg-SVM		Reg-SVM										
سناریوی ۴	سناریوی ۳	سناریوی ۲	سناریوی ۱	GT-SVM	SVM	سناریوی ۴	سناریوی ۳	سناریوی ۲	سناریوی ۱	GT-SVM	SVM	
۰٫۹۸	۰٫۹۹	۰٫۹۸	۰٫۹۸	۰٫۹۸	۰٫۹۶	۰٫۹۹	۰٫۹۹	۰٫۹۹	۰٫۹۹	۰٫۹۹	۰٫۹۹	R <sup>2</sup>
۱٫۸۶	۱٫۰۲	۱٫۹۹	۱٫۹۵	۰٫۹۶	۲٫۸۸	۰٫۷۳	۰٫۶	۰٫۸۸	۰٫۶۳	۰٫۷۶	۰٫۶	RMSE

جدول ۷. مقایسه ماکزیمم مقادیر رسوب معلق پیش‌بینی شده توسط مدل‌های SVM، GT-SVM، و Reg-SVM

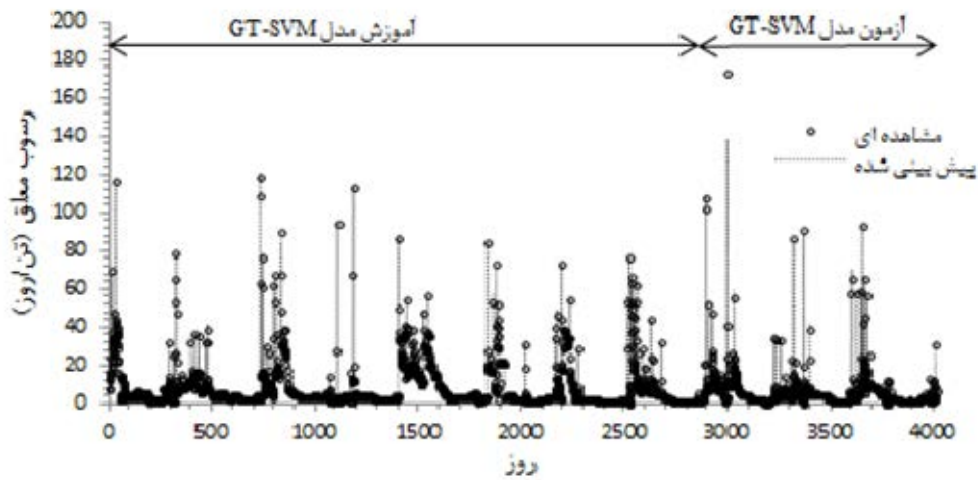
رسوب معلق بیشتر از ۶۰ تن در روز	SVM	Reg-SVM	GT-SVM	خطای نسبی %		
				SVM	Reg-SVM	GT-SVM
۱۷۲,۶۹	۱۴۲,۳۴	۲۴۸,۳۶	۱۳۸,۳۲	-۱۷,۵۸	۴۳,۸۲	-۱۹,۹
۸۶,۷۸	۱۱۱,۵۳	۱۱۵,۷۴	۸۵,۵۷	۲۸,۵۲	۳۳,۳۶	-۱,۴
۹۰,۵۱	۱۲۱,۵۶	۱۲۰,۲۹	۸۹,۲۵	۳۴,۳۱	۳۲,۹	-۱,۳۹
۶۵,۳	۹۲,۵۳	۹۱,۲۴	۶۹,۳	۴۱,۷	۳۹,۷۱	۶,۱۲
۹۳,۱۴	۱۲۲,۲۵	۱۲۰,۵۱	۹۳,۰۲	۳۱,۲۵	۲۹,۳۹	-۰,۱۲
۶۵,۳	۸۸,۷۸	۸۷,۸۱	۶۶,۳۵	۳۵,۹۵	۳۴,۴۶	۱,۶۱

روز را به ترتیب با خطای نسبی معادل ۲۹,۳۹ درصد و ۳۱,۲۵ درصد بیشتر از مقدار واقعی پیش‌بینی می‌کنند، مدل GT-SVM این مقدار رسوب معلق را با حداقل خطای نسبی برابر با ۰,۱۲ درصد کمتر از مقدار واقعی پیش‌بینی می‌کند. به عبارت دیگر، در پیش‌بینی مقادیر پیک رسوب معلق، مدل GT-SVM از توانایی بالاتری نسبت به مدل‌های Reg-SVM و SVM برخوردار است و تخمین دقیق‌تری از میزان ماکزیمم رسوب معلق را ارائه می‌دهد. کل حجم رسوب معلق واقعی در دوره آزمون برابر با ۵۹۲۸,۶۴ تن در دوره آماري است. مدل‌های GT-SVM، Reg-SVM، و SVM این مقدار را به ترتیب برابر با ۵۹۲۷,۴۹ تن در روز، ۹۲۱۱,۰۱ تن در روز، و ۹۰۷۷,۴۶ تن در روز پیش‌بینی کرده‌اند. به عبارتی، حجم رسوب معلق پیش‌بینی شده با استفاده از مدل GT-SVM معادل ۰,۰۲ درصد کمتر از مقدار واقعی است، در صورتی که مدل‌های Reg-SVM و SVM حجم رسوب معلق در دوره آزمون را به ترتیب ۵۵,۳۷ درصد و ۵۳,۱۱ درصد بیشتر از حجم رسوب واقعی پیش‌بینی کرده‌اند. در نتیجه، مدل GT-SVM نسبت به سایر مدل‌ها تخمین دقیق‌تری از حجم رسوب معلق را در دوره آزمون ارائه کرده است. شکل ۵ منحنی رسوب پیش‌بینی شده با استفاده از مدل GT-SVM در مرحله آموزش و آزمون را نشان می‌دهد.

همچنین، جدول ۷ برخی از مقادیر ماکزیمم مقدار رسوب واقعی (بیشتر از ۶۰ تن در روز) و مقادیر رسوب معلق پیش‌بینی شده متناظر با آن در دوره آزمون با استفاده از مدل‌های SVM، GT-SVM، و Reg-SVM را نشان می‌دهد و خطای نسبی میان مقادیر واقعی و مقادیر پیش‌بینی شده بر اساس رابطه ۱۰ محاسبه شده است.

نتایج جدول ۷ نشان می‌دهد که مقادیر پیش‌بینی شده با استفاده از مدل GT-SVM نسبت به مقادیر رسوب پیش‌بینی شده توسط مدل Reg-SVM (تحت سناریوی ۳) و مدل SVM، که هیچ‌گونه پیش‌پردازی بر روی آن صورت نگرفته است، به مقادیر واقعی متناظر آن نزدیک‌تر است. در شرایطی که مدل GT-SVM حداکثر مقدار رسوب واقعی برابر با ۱۷۲,۶۹ تن در روز را برابر با ۱۳۸,۳۲ تن در روز با خطای نسبی ۱۹,۹ درصد کمتر از مقدار واقعی پیش‌بینی می‌کند، مدل Reg-SVM و مدل SVM این مقدار را به ترتیب برابر با برابر ۲۴۸,۳۶ تن در روز و ۱۴۲,۳۴ تن در روز با خطای نسبی به ترتیب ۴۳,۸۲ درصد و ۱۷,۵۸ درصد بیشتر و کمتر از مقدار واقعی پیش‌بینی می‌کنند.

همچنین، در شرایطی که مدل‌های Reg-SVM و SVM مقدار رسوب معلق واقعی برابر با ۹۳,۴ تن در



شکل ۵. رسوب معلق واقعی و پیش‌بینی‌شده در مرحله آموزش و آزمون مدل GT-SVM

جدول ۸. نتایج حاصل از سناریوهای مختلف برای تعیین طول دوره آموزش مدل

سناریو	طول دوره آموزش	آموزش	
		RMSE (تن در روز)	آزمون
۱	۱۰۰۰	۰٫۹۲	۱٫۲۱
۲	۱۵۰۰	۰٫۸۸	۰٫۹۲
۳	۲۰۰۰	۰٫۸	۰٫۹۷
۴	۲۲۵۰	۰٫۷۸	۰٫۹۷
۵	۲۵۰۰	۰٫۸۲	۰٫۸۵
۶	۲۷۵۰	۰٫۸۵	۱٫۰۳
۷	۲۹۰۰	۰٫۷۵	۰٫۸۳
۸	۳۰۰۰	۰٫۸۲	۰٫۸۱
۹	۳۲۵۰	۰٫۹۱	۱٫۰۵
۱۰	۳۵۰۰	۰٫۹۳	۱٫۰۹

این نکته لازم است که در طی ۱۰ سناریوی تعریف‌شده طول دوره آزمون مدل برابر با تعداد ثابت ۵۱۴ داده در نظر گرفته شده است تا تأثیر افزایش طول دوره آموزش بر عملکرد مدل در دوره آزمون بررسی شود. با توجه به مقادیر RMSE نشان‌داده‌شده در جدول ۸، در طی هر ۱۰ سناریوی تعریف‌شده، مقادیر RMSE دوره آزمون بیشتر از مقادیر RMSE

همچنین، برای بررسی نتایج حاصل از روش M تست، ۱۰ سناریوی مختلف تعریف شده است و بر این اساس ۱۰ مدل GT-SVM تدوین شده است. همان‌طور که جدول ۸ نشان می‌دهد، در طی این سناریوها (از سناریوی ۱ تا ۱۰) طول دوره آموزش مدل از ۱۰۰۰ داده به ۳۵۰۰ داده افزایش یافته و عملکرد مدل در هر سناریو بررسی شده است. ذکر

دوره آموزش است. بر اساس نتایج در دوره آموزش، پس از سناریوی ۱ (بیشتر از ۱۰۰۰ دیتا برای آموزش)، مقادیر RMSE از ۰/۹۲ تن در روز به مقدار ۰/۷۵ تن در روز در سناریوی ۷ کاهش می‌یابد و، پس از آن، با افزایش طول دوره آموزش به ۳۵۰۰ دیتا، مقدار RMSE به ۰/۹۳ تن در روز افزایش می‌یابد. همچنین، نتایج دوره آزمون نشان می‌دهد که، متناظر با افزایش طول دوره آموزش مدل، پس از سناریوی ۸ (بیشتر از ۳۰۰۰ دیتا برای آموزش مدل)، مقدار RMSE دوره آزمون افزایش می‌یابد. دقت در این نتایج نشان می‌دهد که کمترین RMSE در طول دوره آزمون مدل بین سناریوی ۷ (۲۹۰۰ داده برای آموزش) تا سناریوی ۸ (۳۰۰۰ داده برای آموزش) اتفاق می‌افتد که با نتایج M تست (۲۹۵۰ داده برای آموزش) مطابقت دارد.

#### ۴. بحث و نتیجه‌گیری

در این مطالعه از یک مدل ماشین بردار پشتیبان برای پیش‌بینی روزانه حجم رسوب معلق رودخانه دوبرج، واقع در استان ایلام، استفاده شد. بهترین ترکیب از متغیرهای ورودی با استفاده از تکنیک نوین آزمون گاما مشخص شد و صحت نتایج آن با روش متداول همبستگی مقایسه شد. همچنین، به منظور بررسی تأثیر پیش‌پردازش متغیرهای ورودی در مدل پیش‌بینی، نتایج حاصل از پیش‌بینی بر مبنای ترکیبات تعیین‌شده توسط آزمون گاما و روش همبستگی با نتایج مدل ماشین بردار پشتیبان، که هیچ پیش‌پردازی بر روی متغیرهای ورودی آن صورت نگرفته بود، با یکدیگر مقایسه شدند. همچنین، طول دوره آموزش مدل با استفاده از آزمون گاما مشخص و صحت نتایج آن در طی تدوین چندین سناریوی مختلف با طول دوره آموزش متفاوت بررسی شد. نتایج این مطالعه نشان داد که پیش‌پردازش متغیرهای ورودی به روش

آزمون گاما و روش متداول همبستگی باعث بهبود عملکرد مدل SVM شده است. همچنین، نتایج نشان داد در صورتی که ورودی‌های مدل با استفاده از روش آزمون گاما انتخاب شوند، خطای پیش‌بینی کمتری نسبت به انتخاب با روش همبستگی در پیش‌بینی حجم رسوب معلق منطقه دارند. نتیجه مطالعه قبلی صورت‌گرفته در زمینه پیش‌بینی جریان نیز نشان داد که مدل پیش‌بینی بر پایه آزمون گاما در مقایسه با مدل پیش‌بینی بر پایه آنالیز مؤلفه اصلی از قابلیت بالاتری در پیش‌بینی جریان برخوردار است [۱۷]. علاوه بر این، در زمینه تخمین تبخیر و تعرق روزانه با استفاده از مدل شبکه عصبی مصنوعی (ANN) بر مبنای گاما تست نتایج مطالعه قبلی محققان نشان داده است که مدل ANN بر مبنای گاما تست از دقت بالاتری نسبت به مدل ANN در شبیه‌سازی برخوردار است. همچنین، اگرچه برتری دقت مدل GT-ANN نسبت به مدل ANN ناچیز بوده است، توانایی مشخص‌نمودن ترتیب اهمیت پارامترهای ورودی و تعیین تعداد داده معنی‌دار برای آموزش شبکه از برتری‌های روش GT بوده است که با نتایج تحقیق حاضر مطابقت دارد [۶]. نتایج این تحقیق نیز نشان داد در صورتی که ورودی‌های مدل با استفاده از روش آزمون گاما انتخاب شوند، دارای خطای پیش‌بینی کمتری نسبت به انتخاب با روش همبستگی در پیش‌بینی حجم رسوب معلق منطقه‌اند. همچنین، نتایج مدل‌سازی رواناب با استفاده از مدل‌های LLR، NNARX، و ANFIS بر مبنای آزمون گاما نشان داده است که GT در تعیین ترکیب مناسب از متغیرهای ورودی و طول مناسب برای آموزش مدل از عملکرد بهتری نسبت به روش‌های متداول و سنتی برخوردار است که با نتایج این تحقیق مطابقت دارد [۱۹]. همچنین، ترکیبات مختلف از داده‌های ورودی با طول دوره آموزش یکسان و یا

معمولاً با وجود کمبود داده‌ها در یک حوزه آبخیز کارایی مناسبی نشان می‌دهند، عملکرد این مدل‌ها بسیار وابسته به الگوهای مورد استفاده در آموزش آنهاست و اگر رویدادی خارج از محدوده آموزش آنها باشد، عملکرد مدل در پیش‌بینی پدیده مورد نظر بسیار ضعیف می‌شود. اگرچه نتایج این مطالعه نشان داد که تکنیک آزمون گاما از توانایی مناسبی در تعیین ترکیب بهینه از متغیرهای ورودی و طول دوره آموزش مدل برخوردار است، مسلماً مطالعات بسیاری برای کسب آگاهی و شناخت بیشتر این تکنیک به منظور تعیین ترکیب مناسب از متغیرهای ورودی و طول دوره آموزش مدل وجود دارد. می‌توان چنین بیان کرد که آزمون گاما می‌تواند تکنیکی نوین برای پیش‌پردازش متغیرهای ورودی در کنار سایر تکنیک‌های پیش‌پردازش، مانند روش تحلیل مؤلفه اصلی، باشد و از آن برای پیش‌بینی‌های هیدرولوژیکی استفاده کرد.

یکی از مشکلاتی که در به‌دست‌آوردن مدل مناسب برای برآورد رسوب معلق عبوری از محل خروجی حوضه‌های آبخیز وجود دارد طیف وسیع میزان عبور است. همان‌گونه که در این تحقیق مشخص است، از کمتر از یک تن در روز تا حدود ۱۷۳ تن در روز بازه این اعداد است. پیشنهاد می‌شود در مطالعات بعدی از مدل‌های کلاس‌بندی SVM برای دسته‌بندی داده‌ها استفاده شود و این داده‌ها به کلاس‌های مناسبی دسته‌بندی شوند. سپس، برای هر دسته بهترین مدل پیش‌بینی تدوین شود.

ترکیب یکسان از متغیرهای ورودی با طول دوره آموزش متفاوت می‌تواند نتایج متفاوتی داشته باشد [۸]. نتایج مطالعه حاضر نیز نشان داد در صورتی که برای یک ترکیب مشخص مدل بر اساس طول متفاوت از داده‌ها آموزش ببیند، عملکرد مدل در پیش‌بینی تغییر می‌کند، به طوری که افزایش طول دوره آموزش مدل از ۱۰۰۰ داده به ۲۹۰۰ داده میزان خطای مدل را به میزان ۳۲ درصد کاهش داد. همچنین، ترکیبات مختلف تعیین‌شده به روش آزمون گاما و روش سنتی رگرسیون بین ورودی و خروجی، با طول دوره یکسان برابر با ۲۹۵۰، هم در مرحله آموزش و هم در مرحله آزمون نتایج متفاوتی ارائه کردند. در این میان ترکیب تعیین‌شده به روش آزمون گاما در مقایسه با سایر ترکیبات از عملکرد بهتری برخوردار بود. همچنین، در پیش‌بینی مقادیر پیک رسوب معلق، مدل GT-SVM بهترین عملکرد را نسبت به سایر مدل‌ها داشته است. بر اساس نتایج آزمون گاما، ترکیب دبی جریان و بارندگی بدون تأخیر زمانی به همراه دبی جریان و میزان بارندگی با دو تأخیر زمانی بهترین ترکیب برای پیش‌بینی میزان رسوبات معلق در دوره مورد مطالعه برای حوزه آبخیز دویرج است. همچنین، تحلیل نتایج نشان داد که آزمون گاما از توانایی بالایی در تعیین طول دوره آموزش مناسب جهت مدل پیش‌بینی برخوردار است. یکی از مسائلی که در مدل‌سازی با استفاده از تکنیک‌های هوش مصنوعی وجود دارد تعیین ترکیب مناسب از متغیرهای ورودی و طول دوره آموزش مدل است. با اینکه تکنیک‌های هوش مصنوعی

## References

- [1] Ahmadi, A., Han, D., Karamouz, M. and Remesan, R. (2009). Input data selection for solar radiation estimation. *Hydrological processes*, 23, 2754-2764.
- [2] Bray, M. and Han, D. (2004). Identification of support vector machines for runoff modeling. *Journal of Hydroinformatics*, 06.4, 265-280.
- [3] Çimen, M. (2008). Estimation of daily suspended sediment using support vector machines. *Hydrological Sciences Journal*, 53(3), 656-666.
- [4] Dibike, Y.B., Velickov, S., Solomatine, D.P. and Abbott, M. (2001). Model induction with support vector machines: introduction and applications. *J. Comput. Civ. Eng.*, 15, 208-216.
- [5] Gericke, A. and Venohre, M. (2012). Improving the estimation of erosion-related suspended solid yields in mountainous, non-alpine river catchments. *Environmental Modelling & Software*, 37, 30-40.
- [6] Ghabaei Sough, M., Mosaedi, A., Hesam, M. and Hezarjaribi, A. (2010). Evaluation effect of input parameters preprocessing in artificial neural networks (Anns) by using stepwise regression and Gamma Test techniques for fast estimation of daily evapotranspiration. *Journal of Water and Soil*, Vol. 24, No. 3, July-Aug 2010, pp. 610-624.
- [7] Han, D., Chan, L. and Zhu, N. (2007a). Flood forecasting using support vector machines. *J. Hydroinform*, 9(4), 267-276, DOI:10.2166/hydro.2007.027.
- [8] Han, D., Kwong, T. and Li, S. (2007b). Uncertainties in real-time flood forecasting with neural networks. *Hydrol. Process*, 21, 223-228.
- [9] Jamalizadeh, M.R., Moghaddamnia, A., Piri, J., Arbabi, V., Homayounifar, M. and Shahryari, A. (2008). Dust storm prediction using ANNs techniques (A case study: Zabol city). *World academy of science, Engineering and Technology*, 43, 512-520.
- [10] Jie, LC. and Yu, ST. (2011). Suspended sediment load estimate using support vector machines in Kaoping River basin, 978-1-61284-459-6/11 2011 IEEE.
- [11] Kisi, O., Haktanir, T., Ardiclioglu, M., Ozturk, O., Yalcin, E. and Uludag, S. (2009). Adaptive neuro-fuzzy computing technique for suspended sediment estimation. *Advances in Engineering Software*, 40, 438-444.
- [12] Kisi, O. (2009). Evolutionary fuzzy models for river suspended sediment concentration estimation. *Journal of Hydrology*, 372, 68-79.
- [13] Kisi, O. (2010). River suspended sediment concentration modeling using a neural differential evolution approach. *Journal of Hydrology*, 389, 227-235.
- [14] Kisi, O., Ozkan, C. and Akay, B. (2012). Modeling discharge-sediment relationship using neural networks with artificial bee colony algorithm. *Journal of Hydrology*, 428-429, 94-103.
- [15] Moghaddamnia, A., Ghafari, M., Piri, J. and Han, D. (2009). Evaporation estimation using support vector machines technique. *International Journal of Engineering and Applied Sciences*, 5(7), 415-423.
- [16] Noori, R., Abdoli, M.A., Ameri, A. and Jalili-Ghazizade, M. (2009). Prediction of municipal solid waste generation with combination of support vector machine and Principal component analysis: a case study of Mashhad. *Environ. Prog. Sustain. Energy*, 28, 249-258.
- [17] Noori, R., Karbassi, A.R., Moghaddamnia, A., Han, D., Zokaei-Ashtiani, M.H., Farokhnia, A. and Ghafari Gousheh, M. (2011). Assessment of input variables determination on the SVM model performance using PCA, Gamma test and forward selection techniques for monthly stream flow prediction. *Journal of Hydrology*, 401, 177-189.
- [18] Rajaei, T. (2011). Wavelet and ANN combination model for prediction of daily suspended sediment load in rivers. *Science of the Total Environment*, 409, 2917-2928.

- [19] Remesan, R., Shamim, M.A., Han, D. and Mathew, J. (2009). Runoff prediction using an integrated hybrid modelling scheme. *J. Hydrol*, 372, 48-60.
- [20] Stefansson, A., Koncar, N. and Jones, A.J. (1997). A note on the gamma test. *Neural Computing & Application*, 5, 131-133.
- [21] Vapnik, V. (1998). *Statistical Learning Theory*, Wiley, New York.
- [22] Vapnik, V. (1995). *The Nature of Statistical Learning Theory*, Springer, New York.
- [23] 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.