

## ارزیابی قابلیت ماشین بردار پشتیبان در پیش‌بینی مونوکسید کربن

صابر معظمی<sup>۱</sup>، روح‌اله نوری<sup>۲\*</sup>، محمد سلیمیان<sup>۳</sup>، محمودرضا مؤمنی<sup>۳</sup>، محمدرضا وصالی ناصح<sup>۴</sup>

- ۱- استادیار، گروه مهندسی عمران، مرکز تحقیقات علوم زیست‌محیطی، واحد اسلامشهر، دانشگاه آزاد اسلامی
- ۲- استادیار، دانشکده تحصیلات تکمیلی محیط‌زیست، دانشگاه تهران
- ۳- کارشناس ارشد، دانشکده تحصیلات تکمیلی محیط‌زیست، دانشگاه تهران
- ۴- استادیار، گروه مهندسی عمران، دانشگاه اراک

\*noor@ut.ac.ir

تاریخ پذیرش: [۹۵/۰۳/۱۲]

تاریخ دریافت: [۹۴/۱۰/۲۸]

### چکیده -

هدف اصلی این پژوهش تعیین عدم قطعیت ماشین بردار پشتیبان (SVM) برای تخمین مونوکسید کربن (CO) در ایستگاه پایش قلهک، در شمال تهران، است. برای تعیین عدم قطعیت مدل SVM، با تغییر الگوهای انتخابی برای کالیبراسیون، به دفعات زیاد واسنجی مدل انجام شد. سپس با استفاده از فاکتور عرض باند ( $d$ -factor) و درصد داده‌های اندازه‌گیری قرار گرفته در باند ۹۵ درصد عدم قطعیت مدل (95PPU)، عملکرد SVM در فرآیند پیش‌بینی CO ارزیابی شد. نتایج این پژوهش بیانگر عملکرد مناسب مدل SVM در پیش‌بینی غلظت روزانه CO در شهر تهران بود به گونه‌ای که از ضریب تعیین ( $R^2$ ) معادل ۰/۸۹ و ۰/۸۸ به ترتیب در مراحل کالیبراسیون و درستی‌آزمایی برخوردار بود. همچنین یافته‌های تحلیل عدم قطعیت مدل SVM مشخص نمود که مقدار  $d$ -factor و درصد داده‌های اندازه‌گیری محصور شده به وسیله 95PPU برای این مدل در مرحله درستی‌آزمایی به ترتیب معادل ۰/۷۴ و ۷۶ درصد است.

**واژگان کلیدی:** ماشین بردار پشتیبان، آلودگی هوا، تحلیل عدم قطعیت، تهران، مونوکسید کربن.

### ۱- مقدمه

قرار دهد. در این راستا بررسی و ارائه الگوهای جدید در تعیین عدم قطعیت مدل‌های پیش‌بینی کیفی هوا از دیدگاه بهداشتی، به دلیل تأثیرگذاری بر سلامت افراد بسیار زیاد در معرض آلودگی، از اهمیت ویژه‌ای برخوردار است. با وجود مطالعات کافی برای تحلیل عدم قطعیت مدل‌های عددی آلودگی هوا [1-3]، متأسفانه مراجع بسیار اندکی در این زمینه برای مدل‌های هوشمند در دسترس است. مدل‌های هوشمند از قابلیت بالایی در پیش‌بینی آلاینده‌های هوا برخوردارند و در بسیاری از موارد دارای مزیت‌هایی بر مدل‌های عددی است. این مزیت‌ها بیشتر شامل

اولین گام برای مقابله با معضل آلودگی هوا اطلاع دقیق از وضعیت کیفی آن با استفاده از مدل‌ها به منظور تدوین راهبردهای لازم است. در سال‌های اخیر مطالعات پیش‌بینی آلودگی هوا بیشتر با استفاده از مدل‌های عددی و همچنین مدل‌های هوشمند انجام شده است. به هر حال اگرچه تلاش‌های زیادی برای آشکارسازی کیفیت هوا به وسیله مدل‌های پیش‌بینی صورت گرفته، اما وجود عدم قطعیت در خروجی این مدل‌ها می‌تواند نتایج آنها را به شدت تحت تأثیر

دی‌اکسید سولفور ( $SO_2$ ) و ذرات معلق به کار گرفته و نتایج مدل مناسب ارزیابی شد [10]. پژوهشی دیگر به بررسی عملکرد مدل SVM برای پیش‌بینی ازن ( $O_3$ ) در لایه تروپوسفری اتمسفر در مرز بین امریکا و مکزیک پرداخت [11]. همچنین پژوهشگران دیگر با استفاده از شرایط اتمسفری به عنوان تنها ورودی به مدل SVM، نتایج مطلوبی برای پیش‌بینی  $O_3$  به دست آوردند [12]. برای شهر تهران نیز عملکرد مدل SVM در پیش‌بینی آلاینده‌ها مثبت گزارش شده است [13]. استفاده از مدل SVM برای پیش‌بینی آلاینده‌های هوا در مقایسه با دیگر مدل‌های هوشمند نیز بیانگر عملکرد برتر این مدل بوده است [14]. پژوهش دیگر نیز از مدل SVM برای پیش‌بینی ذرات معلق با قطر کمتر یا مساوی ۱۰ میکرومتر ( $PM_{10}$ ) در پکن استفاده کرده است [15]. به هر حال ساختار تئوریک پیچیده مدل SVM در مقایسه با مدل‌های ANN و ANFIS به گونه‌ایست که تحلیل عدم قطعیت این مدل را عملاً مشکل می‌نماید زیرا ساختار مدل در توسعه روشی برای تحلیل عدم قطعیت آن لازمست لحاظ شود. به همین دلیل تاکنون سند علمی مناسبی برای تحلیل عدم قطعیت مدل SVM در مراجع علمی ارائه نشده است. بنابراین با توجه به مطالب مذکور هدف اصلی این پژوهش ارائه روش‌شناسی مناسبی برای تحلیل عدم قطعیت مدل SVM در فرآیند پیش‌بینی آلاینده‌های CO است. برای این منظور محدوده مطالعه شده این پژوهش شهر تهران انتخاب شد و با استفاده از اطلاعات آلودگی هوا و هواشناسی، این روش‌شناسی ارزیابی شد.

## ۲- روش تحقیق

### ۲-۱- محدوده مطالعه شده و اطلاعات مسأله

حدود ۹۰ درصد وزنی آلاینده‌های هوا در تهران به وسایل نقلیه و ۱۰ درصد به دیگر منابع اختصاص دارد [5]. این پژوهش در ادامه مقالات چاپ شده پژوهشگر صورت گرفته است [4-6]. بنابراین داده‌های آلودگی و اطلاعات هواشناسی

عدم نیاز به اطلاعات انتشار و ضرایب انتشار آلاینده‌ها، ساختار ساده‌تر آنها و همچنین انعطاف‌پذیری بیشتر این مدل‌ها در مواجهه با کمبود اطلاعات در مقایسه با مدل‌های عددی است [4-5]. این مزیت‌ها در کنار قابلیت بالای این مدل‌ها در پیش‌بینی پدیده‌های غیرخطی و پیچیده دلیل اصلی استفاده روزافزون از آنها در مطالعات پیش‌بینی آلاینده‌های هوا بوده است. به هر حال عدم وجود راهکاری مناسب برای تحلیل عدم قطعیت این مدل‌ها بزرگترین چالش پیش‌رو در استفاده از نتایج آنها به منظور تصمیم‌گیری‌های مدیریتی است. تنها مطالعه موجود در این راستا مربوط به تحلیل عدم قطعیت مدل‌های هوشمند شبکه عصبی مصنوعی<sup>۱</sup> (ANN) و سیستم استنتاج نروفازی تطبیقی<sup>۲</sup> (ANFIS) برای پیش‌بینی غلظت روزانه مونوکسید کربن<sup>۳</sup> (CO) در شهر تهران است [6] از دیگر مدل‌های هوش مصنوعی می‌توان به ماشین بردار پشتیبان<sup>۴</sup> (SVM) اشاره نمود که به دلیل احتمال بیشتر در دستیابی به جواب بهینه نسبت به دو مدل ANN و ANFIS در سالیان اخیر مورد توجه متخصصین آلودگی هوا قرار گرفته است [7-8]. در مقایسه با مدل‌های ANN و ANFIS، تعداد پارامترهای کمتر در مدل SVM برای بهینه‌سازی و استفاده از برنامه کمینه‌سازی ریسک تجربی نسبت به کمینه‌سازی ریسک ساختاری این امکان را برای SVM فراهم می‌نماید تا از عملکرد معادل یا حتی بهتری نسبت به دو مدل دیگر برخوردار باشد. این مزیت دلیل اصلی رویکرد پژوهشگران آلودگی هوا به مدل SVM در دهه اخیر بوده است [9] در پژوهشی امکان استفاده از مدل SVM برای پیش‌بینی آلاینده‌های هوا در شهر هنک‌کنگ چین بررسی و نتایج این مدل در مقایسه با مدل ANN برتر گزارش شد [9]. همچنین در پژوهش دیگری مدل ترکیبی SVM و تبدیل موجک برای شبیه‌سازی غلظت دی‌اکسید نیتروژن ( $NO_2$ )، CO،

1-Artificial Neural Network (ANN)

2-Adaptive Neuro-Fuzzy Inference System (ANFIS)

3-Carbon Monoxide (CO)

4-Support Vector Machine (SVM)

در معادلات بالا  $C$  عددی صحیح مثبت است که عامل تعیین جرمه در هنگام رخ دادن خطای آموزش مدل است.  $N$  تعداد نمونه‌ها و دو پارامتر  $\xi_i$  و  $\xi_i^*$  متغیرهای کمبود<sup>۳</sup> هستند که حد بالا و پایین خطای آموزش مرتبط با مقدار خطای مجاز  $\epsilon$  را مشخص می‌کنند. اطلاعات بیشتر در مورد مدل SVM در مراجع [20-22] موجود است.

### ۳-۲- تحلیل عدم قطعیت مدل SVM

با توجه به آنچه در بالا در مورد مدل SVM مطرح شد، ضروری است تابع خطای مربوط به داده‌های ورودی مدل بهینه شود. بنابراین لازم است تا کارکرد مدل تحت الگوهای واسنجی متفاوت ارزیابی شده تا عدم قطعیت مدل حاصل از تغییرات ایجاد شده در داده‌های ورودی بررسی شود. بدین منظور، یک الگوی واسنجی طراحی و به دنبال آن داده‌های خروجی به وسیله مدل محاسبه شد. لازم به گفتن است که این فرآیند باید به تعداد زیادی تکرار شود. در این پژوهش، مدل SVM به وسیله درصدی از داده‌ها واسنجی شد و فرآیند نمونه‌گیری داده‌ها برای واسنجی این مدل به تعداد مناسب مورد تکرار قرار گرفت (در این پژوهش، ۱۰۰۰ مرتبه). در نهایت پارامترهای مدل ۱۰۰۰ مرتبه محاسبه شد. بدین روی، دامنه‌ای از داده‌های خروجی مربوط به عدم قطعیت مدل واسنجی شده SVM با بکارگیری این روش محاسباتی حجیم تعیین شد. به منظور ارزیابی عدم قطعیت SVM درصد داده‌های اندازه‌گیری شده به وسیله ۹۵ درصد پیش‌بینی‌های قرار گرفته در باند اطمینان (95PPU) مطابق معادله ۳ محاسبه شد. برای این منظور نیاز به تعیین ۲/۵ درصد حد پایین پیش‌بینی‌ها ( $X_L$ ) و ۹۷/۵ درصد حد بالای آن ( $X_U$ ) حاصل از توزیع نرمال تجربی ناشی از ۱۰۰۰ مرتبه تکرار فرآیند واسنجی SVM است. همچنین برای ارزیابی عرض باند اطمینان معیار  $d$ -factor که بیانگر عرض میانگین باند فاصله اطمینان می‌باشد مطابق فرمول ۴ مورد استفاده قرار گرفت [23].

سال‌های ۲۰۰۴ و ۲۰۰۵ از ایستگاه قل‌هک در شمال تهران به منظور واسنجی و درستی‌آزمایی مدل SVM به کار گرفته شد. از این رو برای پیش‌بینی غلظت روزانه CO در آینده، میانگین حسابی روزانه شش آلاینده هوا شامل PM، مجموع ترکیبات هیدروکربن (THC)، اکسیدهای نیتروژن ( $\text{NO}_x$ )، متان ( $\text{CH}_4$ )،  $\text{SO}_2$  و  $\text{O}_3$  و همچنین شش متغیر هواشناسی شامل فشار (Press)، دمای هوا (Temp)، جهت وزش باد (WD)، سرعت وزش باد (WS) و رطوبت نسبی (Hum) استفاده شده است.

### ۲-۲- ماشین بردار پشتیبان

در یک مدل رگرسیونی SVM، تابعی مرتبط با متغیر وابسته  $Y$  که خود تابعی از چند متغیر مستقل  $x$  است برآورد می‌شود. در این نوع مدل فرض می‌شود که رابطه میان متغیرهای مستقل و وابسته توسط یک تابع جبری مانند  $f(x)$  (  $f(x) = W^T \cdot \phi(x) + b$  ) به علاوه مقداری اغتشاش<sup>۱</sup> مشخص شود (  $Y = f(x) + \text{noise}$  ) [16-18]. در کتب مرجع مرتبط با SVM اغتشاش به عنوان خطای مجاز ( $\epsilon$ ) تعریف شده است. چنانچه  $w$  (بردار ضرایب) و  $b$  (ثابت) پارامترهای تابع رگرسیونی و  $\phi$  نیز تابع کرنل<sup>۲</sup> باشد، آنگاه هدف پیدا کردن یک فرم تابعی برای  $f(x)$  است. این مهم با آموزش SVM به وسیله مجموعه‌ای از نمونه‌ها محقق می‌شود. این پروسه شامل بهینه‌سازی متوالی یک تابع خطا معادله ۱ در SVM با در نظر گرفتن شرایط مندرج در معادله ۲ است [19]

$$\frac{1}{2} W^T \cdot W + C \sum_{i=1}^N \xi_i + C \sum_{i=1}^N \xi_i^* \quad (1)$$

$$W^T \cdot \phi(x_i) + b - y_i \leq \epsilon + \xi_i^* \quad (2)$$

$$Y_i - W^T \cdot \phi(x_i) - b \leq \epsilon + \xi_i$$

$$i = 1, \dots, N, \quad \xi_i, \xi_i^* \geq 0$$

1-Noise

2-Kernel Function

3-Slack Variables

داده‌های جایگزین شامل ۲۰۰ داده برای نمونه‌گیری استفاده شد. بنابراین، الگوریتم نمونه‌گیری طراحی شده احتمال تغییر ۲۰ درصدی داده‌های واسنجی را به وسیله جابه‌جایی تصادفی آنها با داده‌های جایگزین فراهم آورد. سپس پارامترهای  $C$ ،  $\epsilon$  و  $\gamma$  با تکرار ۱۰۰۰ مرتبه واسنجی مدل با الگوهای انتخابی کالیبراسیون برای مدل SVM به دست آمد و در پی آن نتایج برای هر تکرار ذخیره شد. در مرحله بعد با محاسبه عرض باند اطمینان، عدم قطعیت مدل SVM برای برآورد غلظت روزانه CO در طول فرآیند واسنجی و صحت‌سنجی مطابق شکل (۴) محاسبه شد. همچنین علاوه بر ارائه نتایج عدم قطعیت مدل SVM در فرآیند تخمین روزانه CO، معیارهای آماری نتایج شامل آماره  $d$ -factor و مقادیر محدود شده به وسیله 95PPU در مراحل واسنجی و درستی‌آزمایی مدل SVM در جدول (۱) آمده است.

مطابق نتایج به دست آمده مشخص است که نتایج پیش‌بینی شده به وسیله مدل از هماهنگی مناسبی با اطلاعات اندازه‌گیری شده برخوردار است. از شکل (۱) مشخص است که در مرحله واسنجی مدل تقریباً نتایج پیش‌بینی شده هماهنگ بر غلظت‌های اندازه‌گیری شده در ایستگاه قلهدک است هرچند که مدل در پیش‌بینی برخی مقادیر شامل غلظت‌های بالای ثبت شده CO در این ایستگاه اندکی با خطا همراه است. همچنین ارزیابی این مدل در مرحله درستی‌آزمایی نیز بیانگر دقت قابل قبول این مدل است.

$$\text{Bracketed by 95 PPU} = \frac{1}{n} \text{count} (Q | X_L \leq Q \leq X_U) \times 100 \quad (3)$$

$$d\text{-factor} = \frac{\overline{d_x}}{\sigma_x} \quad (4)$$

در رابطه ۴،  $\sigma_x$  انحراف معیار متغیر  $x$  و  $\overline{d_x}$  میانگین فاصله بین کران بالا و پایین بدست آمده از رابطه زیر است [27-24].

$$\overline{d_x} = \frac{1}{k} \sum_{i=1}^k (X_{U_i} - X_{L_i}) \quad (5)$$

### ۳- یافته‌های تحقیق

تابع RBF به عنوان تابع کرنل در این پژوهش انتخاب شد و با توجه به اطلاعات موجود برای پیش‌بینی CO، مدل SVM توسعه داده شد. با توجه به انتخاب تابع کرنل RBF، ضروری است تا دو پارامتر مدل SVM ( $C$  و  $\epsilon$ ) و همچنین پارامتر تابع کرنل RBF ( $\gamma$ ) بهینه شوند. برای یافتن مقادیر بهینه این پارامترها از روش بهینه‌سازی الگوریتم جستجوی شبکه دو مرحله‌ای<sup>۱</sup> استفاده شد. برای این منظور در مرحله اول بهینه‌سازی، ابعاد شبکه انتخابی بزرگ انتخاب شدند و بعد از یافتن محدوده جواب بهینه، ابعاد کوچک‌تری برای افزایش دقت و یافتن جواب بهینه مورد استفاده قرار گرفت. با یافتن مقادیر بهینه سه پارامتر مذکور، عملاً مدل پیش‌بینی CO توسعه یافت که نتایج مراحل واسنجی و درستی‌آزمایی آن در شکل‌های (۱ تا ۳) آمده است.

برای انتخاب الگوهای واسنجی مدل نیز یک مجموعه

شکل ۱. روند زمانی تغییرات مقادیر مشاهداتی در مقابل مقادیر پیش‌بینی شده CO توسط مدل SVM

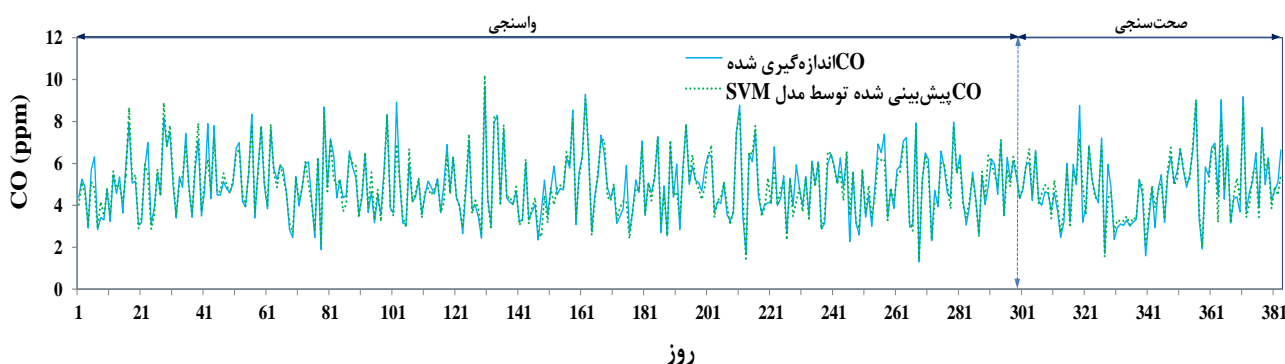


Fig. 1. Predicted vs observed CO (ppm) by SVM model

دیگر و در بیان کمی می‌توان بیان نمود که مدل برای مقدار  $R^2$  معادل واحد عملاً هیچ‌گونه خطایی ندارد و دارای بهترین عملکرد است. همچنین به ازای  $R^2$  معادل صفر مدل بدترین عملکرد خود را داشته و این بدان معناست که عملاً هیچ‌گونه ارتباطی بین غلظت CO با پارامترهای ورودی به مدل (متغیرهای آلودگی هوا و هواشناسی) وجود ندارد. بنابراین با توجه به مطالب مذکور می‌توان دریافت که مدل SVM با ضریب  $R^2$  معادل ۰/۸۹ و ۰/۸۸ به ترتیب برای مراحل واسنجی و درستی‌آزمایی از عملکرد مطلوبی در فرآیند پیش‌بینی غلظت روزانه CO در ایستگاه قلحک برخوردار بوده است.

جدول ۱. نتایج تعیین عدم قطعیت مدل SVM در پیش‌بینی روزانه CO در

شهر تهران		
Step	Statistics	SVM Model
Calibration	d-factor	0.71
	Bracketed by 95PPU (%)	78
Verification	d-factor	0.74
	Bracketed by 95PPU (%)	76

Table 1. Uncertainty results of calibration and testing steps for SVM model

شکل ۴. ۹۵ درصد سطح اطمینان به دست آمده برای مدل SVM در فرآیند

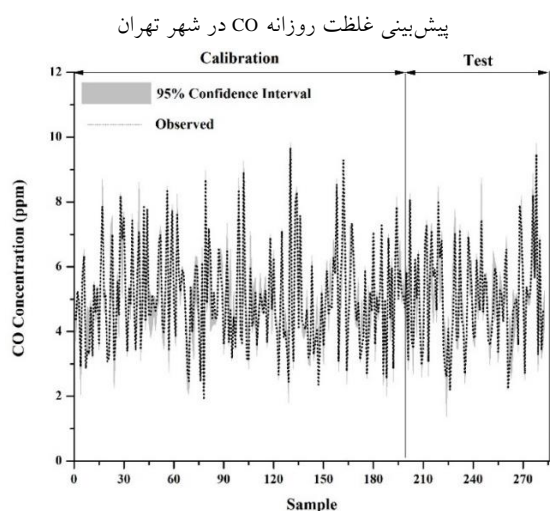


Fig. 4. 95% confidence intervals for the estimates of daily CO concentration during the calibrating step using SVM model

نتایج به دست آمده از تعیین عدم قطعیت مدل SVM نیز بیانگر عدم قطعیت اندک این مدل است. همان‌گونه که در

شکل ۲. مقادیر مشاهده شده در مقابل مقادیر پیش‌بینی شده CO در مرحله

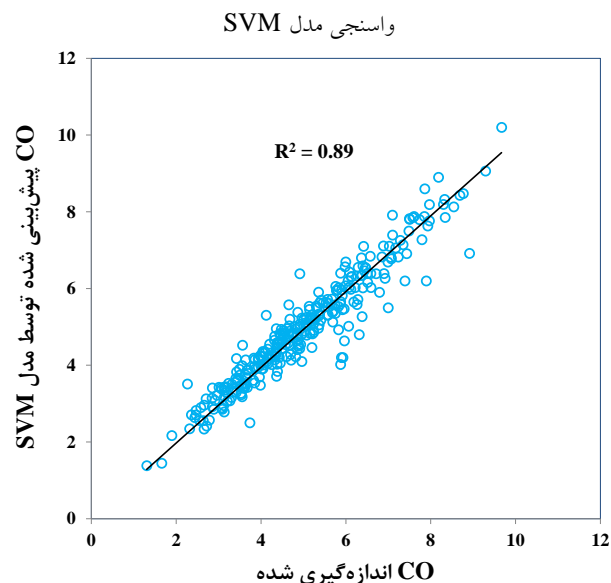


Fig. 2. Scatter plot of observed and predicted CO (ppm) by SVM model in calibration step

شکل ۳. مقادیر مشاهده شده در مقابل مقادیر پیش‌بینی شده CO در مرحله

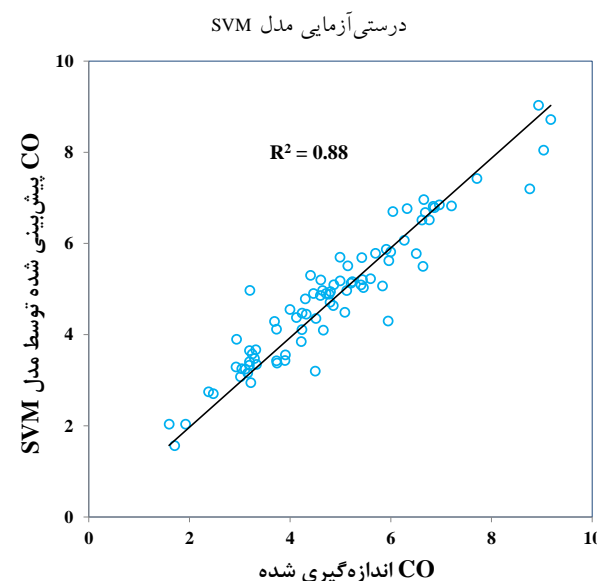


Fig. 3. Scatter plot of observed and predicted CO (ppm) by SVM model in testing step

واقعیت‌های مذکور در شکل‌های (۲ و ۳) نیز به شکل دیگر مشاهده می‌شود. مطابق این شکل‌ها بهترین حالت زمانی است که نتایج به دست آمده از مدل در مقابل مقادیر اندازه‌گیری شده در مراحل واسنجی و درستی‌آزمایی روی خطی با زاویه ۴۵ درجه قرار گیرند که در این صورت خطای مدل صفر شده و نتایج پیش‌بینی شده نیز عیناً منطبق بر اطلاعات ثبت شده CO در ایستگاه قلحک می‌شود. به عبارتی

تهران برخوردار بود. در پایان قابل ذکر است که روش‌شناسی ارائه شده در این پژوهش برای تحلیل عدم قطعیت SVM کلی بوده و قابل تعمیم برای پیش‌بینی سایر پدیده‌ها در دیگر زمینه‌های علوم و مهندسی است.

## ۵- تشکر و قدردانی

این پژوهش برگرفته از طرح پژوهشی "تعیین عدم قطعیت‌های مدل ماشین بردار پشتیبان در پیش‌بینی مونوکسیدکربن در هوای شهر تهران" است که با حمایت مالی دانشگاه آزاد اسلامی واحد اسلامشهر به اجرا درآمده است، بدینوسیله از آن واحد دانشگاهی، تقدیر و تشکر به عمل می‌آید.

## References

## ۶- مراجع

- [1] Hanna S. R. 1998 Air quality model evaluation and uncertainty. *Journal Air Pollution Control Association*, 38(4), 406-412.
- [2] Kioutsioukis L., Tarantola S., Saltelli A. & Gatelli D. 2004 Uncertainty and global sensitivity analysis of road transport emission estimates. *Atmospheric Environment*, 38(38), 6609-6620.
- [3] Özkaynak H., Frey H. C., Burke J. & Pinder R. W. 2009 Analysis of coupled model uncertainties in source-to-dose modeling of human exposures to ambient air pollution: A PM<sub>2.5</sub> case study. *Atmospheric Environment*, 43(9), 1641-1649.
- [4] Noori R., Hoshyaripour G., Ashrafi K. & Rasti O. 2013 Introducing an appropriate model using support vector machine for predicting carbon monoxide daily concentration in Tehran atmosphere. *Iranian Journal of Health and Environment*, 6(1), 1-10 (In Persian).
- [5] Noori R., Ashrafi K. & Ajdarpour A. 2008 Comparison of ANN and PCA based multivariate linear regression applied to predict the daily average concentration of CO: A case study of Tehran. *Journal of the Earth and Space Physics*, 34(1), 135-152 (In Persian).
- [6] Noori R., Hoshyaripour G., Ashrafi K. & Araabi B. N. 2010 Uncertainty analysis of developed ANN and ANFIS models in prediction of carbon monoxide daily concentration. *Atmospheric Environment*, 44(4), 476-482.
- [7] Lu W. Z. & Wang W. 2005 Potential assessment of the "support vector machine" method in forecasting ambient air pollutant trends. *Chemosphere*, 59(5), 693-701.
- [8] Lu W. Z. & Wang D. 2008 Ground-level ozone prediction by support vector machine approach with a

جدول (۱) دیده می‌شود، بخش قابل توجهی از پیش‌بینی‌های مدل SVM (بیش از ۷۸ درصد و ۷۶ درصد به ترتیب در مراحل واسنجی و درستی‌آزمایی در محدوده 95PPU قرار می‌گیرند. همچنین آماره  $d$ -factor برای مدل SVM در هر دو مرحله واسنجی و درستی‌آزمایی مدل کمتر از ۱ است. این حقیقت بیانگر عملکرد مناسب مدل SVM و عدم قطعیت قابل قبول آن برای پیش‌بینی غلظت روزانه CO در شهر تهران است. به طورکلی بهترین عملکرد مدل بر مبنای شاخص‌های عدم قطعیت برای  $d$ -factor معادل صفر و برای مقادیر محدود شده به وسیله 95PPU معادل ۱۰۰ درصد است که در این صورت مدل عملاً هیچ‌گونه عدم قطعیتی ناشی از تغییر در داده‌های انتخابی برای واسنجی از خود نشان نمی‌دهد. ولی به هر حال باید توجه نمود که مدل‌هایی مانند SVM داده‌محور بوده و بر مبنای داده‌های ورودی واسنجی می‌شوند. بنابراین تغییر داده‌های واسنجی در این مدل همیشه همراه با مقداری عدم قطعیت است.

## ۴- نتیجه‌گیری

پیش‌بینی‌های کوتاه مدت آلودگی هوای شهری به کمک روش‌های هوش مصنوعی برای مسئولین و نهادهای شهری و به منظور اتخاذ راهبردهای کنترلی مناسب، ضروری می‌باشد. با این حال، نتایج مدل‌های هوش مصنوعی قطعی نبوده و تحت تأثیر حاشیه خطای قابل توجهی قرار می‌گیرد. بنابراین محاسبات عدم قطعیت مدل‌های هوشمند به خصوص در بخش مدیریت راهبردی کنترل آلودگی هوا می‌تواند بسیار حیاتی باشد. از این رو هدف اصلی این پژوهش ارائه روشی مناسب جهت محاسبه عدم قطعیت مدل SVM برای پیش‌بینی غلظت روزانه CO قرار داده شد. یافته‌های این پژوهش مشخص نمود که روش‌شناسی ارائه شده در این پژوهش به خوبی قادر به محاسبه عدم قطعیت مدل SVM برای پیش‌بینی غلظت روزانه CO بود. همچنین بر اساس شاخص‌های آماری  $d$ -factor و 95PPU، عدم قطعیت مدل SVM برای پیش‌بینی غلظت روزانه CO در تهران قابل قبول بوده به گونه‌ای که این مدل از عدم قطعیت اندکی در پیش‌بینی این شاخص کیفیت هوا در شهر



- [18] Moazami S., Noori R., Amiri B. J., Yeganeh B., Partani S. & Safavi S. 2016 Reliable prediction of carbon monoxide using developed support vector machine. *Atmospheric Pollution Research*, 7(3), 412-418.
- [19] Vapnik V. N. 1998 *Statistical Learning Theory*. Wiley, New York.
- [20] Fletcher R. 1987 *Practical Methods of Optimization*. Wiley, New York.
- [21] Abe S. 2005 *Support Vector Machines for Pattern Classification*. Springer-Verlag, London.
- [22] Noori R., Karbassi A. R., Moghaddamia A., Han D., Zokaei-Ashtiani M. H., Farokhnia A. & Ghaffari-Goushe 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(3), 177-189.
- [23] Abbaspour K. C., Yang J., Maximov I., Siber R., Bogner K., Mieleitner J., Zobrista J. & Srinivasan R. 2007 Modeling hydrology and water quality in the pre alpine/alpine Thur watershed using SWAT. *Journal of Hydrology*, 333(2), 413-430.
- [24] Noori R., Safavi S. & Shahrokni A. A. N. 2013 A reduced-order adaptive neuro-fuzzy inference system model as a software sensor for rapid estimation of fiveday biochemical oxygen demand. *Journal of Hydrology*, 495, 175-185.
- [25] Noori R., Deng Z., Kiaghadi A. & Kachoosangi F. T. 2016 How reliable are ANN, ANFIS, and SVM techniques for predicting longitudinal dispersion coefficient in natural rivers?. *Journal of Hydraulic Engineering*, DOI:10.1061/(ASCE)HY.1943-7900.0001062.
- [26] Dehghani M., Saghafian B., Nasiri Saleh F., Farokhnia A. & Noori R. 2014 Uncertainty analysis of streamflow drought forecast using artificial neural networks and Monte-Carlo simulation. *International Journal of Climatology*, 34(4), 1169-1180.
- [27] Noori R., Ghiasi B., Sheikhan H., Adamowski J. F. 2017 Estimation of the dispersion coefficient in natural rivers using a granular computing model. *Journal of Hydraulic Engineering*, DOI:10.1061/(ASCE)HY.1943-7900.0001276.
- cost-sensitive classification scheme. *Science of the Total Environment*, 395(2), 109-116.
- [9] Lu W. Z., Wang W. J., Fan H. Y., Leung A. Y. T., Xu Z. B. & Lo S. M. 2002 Air pollutant parameter forecasting using support vector machines. *IEEE*, 1, 630-635.
- [10] Osowski S. & Garanty K. 2007 Forecasting of the daily meteorological pollution using wavelets and support vector machine. *Engineering Application of Artificial Intelligent*, 20(6), 745-755.
- [11] Salazar-Ruiz E., Ordieres J. B., Vergara E. P. & Capuz-Rizo S. F. 2008 Development and comparative analysis of tropospheric ozone prediction models using linear and artificial intelligence-based models in Mexicali, Baja California (Mexico) and Calexico, California (US). *Environmental Modelling and Software*, 23(8), 1056-1069.
- [12] Feng Y., Zhang W., Sun D. & Zhang L. 2011 Ozone concentration forecast method based on genetic algorithm optimized back propagation neural networks and support vector machine data classification. *Atmospheric Environment*, 45(11), 1979-1985.
- [13] Yeganeh B., Motlagh M. S. P., Rashidi Y. & Kamalan H. 2012 Prediction of CO concentrations based on a hybrid Partial Least Square and Support Vector Machine model. *Atmospheric Environment*, 55, 357-365.
- [14] Singh K. P., Gupta S. & Rai P. 2013 Identifying pollution sources and predicting urban air quality using ensemble learning methods. *Atmospheric Environment*, 80, 426-437.
- [15] Weizhen H., Zhengqiang L., Yuhuan Z., Hua X., Ying Z., Kaitao L., Donghui L., Peng W. & Yan M 2014 Using support vector regression to predict PM<sub>10</sub> and PM<sub>2.5</sub>. In: *IOP Conference Series: Earth and Environmental Science*, IOP Publishing, 17(1), pp. 012268.
- [16] Noori R., Karbassi A., Ashrafi K., Ardestani M., Mehrdadi N. & Bidhendi G. R. N. 2012 Active and online prediction of BOD<sub>5</sub> in river systems using reduced-order support vector machine. *Environmental Earth Sciences*, 67(1), 141-149.
- [17] Noori R., Yeh H. D., Abbasi M., Kachoosangi F. T. & Moazami S. 2015 Uncertainty analysis of support vector machine for online prediction of five-day biochemical oxygen demand. *Journal of Hydrology*, 527, 833-843.

# Evaluation of Support Vector Machine Performance for Carbon Monoxide Prediction

Saber Moazami<sup>1</sup>, Roohollah Noori<sup>\*2</sup>, Mohammad Salimian<sup>3</sup>, Mahmudreza Momeni<sup>3</sup>,  
Mohammad Reza Vesali Naseh<sup>4</sup>

1- Assist. Prof., Department of Civil Engineering, Environmental Sciences Research Center, Islamshahr Branch, Islamic Azad University, Islamshahr, Iran

2- Assist. Prof., Graduate Faculty of Environment, University of Tehran, Tehran, Iran

3- M.Sc., Graduate Faculty of Environment, University of Tehran, Tehran, Iran

4- Assist. Prof., Department of Civil Engineering, Arak University, Arak, Iran

noor@ut.ac.ir

## Abstract

Carbon monoxide (CO) is one of the main air pollutant parameters in the atmosphere of Tehran, Iran. Generally, it is difficult to predict and control the intensity of CO concentration because it essentially has nonlinear time-varying nature. Environmental control such as controlling the CO concentration level is recently regarded as one of the most important factors in environmental protections. This paper describes forecasting and - more specifically - determination of uncertainty in evaluating the CO concentration during the modeling process using a support vector machine (SVM) technique. Uncertainty of air pollution modeling highly affects the simulation results. In this regards, it is very important to determine the uncertainty of air pollution models due to consequences on the health of people exposed to the pollution. Therefore, this research aims to calibrate, verify, and also determine the uncertainty of SVM in the process of air pollution modeling in the atmosphere of Tehran. To achieve this goal, the SVM model was applied to predict arithmetic average of daily measured CO concentration in the atmosphere of Tehran. In this regard, the SVM model was calibrated and verified using six daily air pollutants including particulate matter (PM), total hydrocarbons (THC), nitrogen oxides (NO<sub>x</sub>), methane (CH<sub>4</sub>), sulfur dioxide (SO<sub>2</sub>) and ozone (O<sub>3</sub>) and also six daily meteorological variables including pressure (Press), temperature (Temp), wind direction (WD), wind speed (WS) and relative humidity (Hum). The data was collected from Gholhak station located in the north of Tehran, Iran, during 2004-2005. Thereafter, the best developed SVM model for predicting CO concentration was chosen based on determination of coefficient ( $R^2$ ). Finally, to determine the SVM uncertainty, the model was run many times with different calibration data. It led to many different results because of the model sensitivity to the selected calibration data. Then, the model uncertainty in the CO prediction process was evaluated using the width of uncertainty band (*d*-factor) and the percentage of measured data bracketed by the 95 percent prediction uncertainties (95PPU). Generally, the results confirmed the strong performance of the SVM model in predicting CO concentration in the atmosphere of Tehran. The predicted average daily CO concentrations by SVM model had a good agreement with the measured ones in the Gholahak air quality monitoring station. It was found that determination of coefficient for calibration and validation of SVM model were equal to 0.89 and 0.88, respectively. Furthermore, the results indicated that the SVM model has an acceptable level of uncertainty in predicting CO concentration. In this case, the level of *d*-factor and the percentage of measured data - bracketed by the 95PPU in the validation step - were 0.74 and 76, respectively. Therefore, obtained results indicated that the SVM model had an acceptable level of uncertainty in prediction of CO concentration. The proposed methodology is general, therefore; it is suggested to be applied for analyzing the uncertainty of SVM model in other fields of science and engineering.

**Keywords:** Support Vector Machine, Air Pollution, Uncertainty Analysis, Tehran, Carbon Monoxide