

# پیشبینی سری زمانی مقدار اُزن تروپوسفری با استفاده از شبکه عصبی مصنوعی و ماشین بردار پشتیبان

وحید مهدی پور<sup>۱</sup>، مهسا معماریان فرد<sup>۲\*</sup>

- ۱- کارشناس ارشد عمران محیط زیست، دانشگاه صنعتی خواجه نصیرالدین طوسی
- ۲- استادیار دانشکده مهندسی عمران، دانشگاه صنعتی خواجه نصیرالدین طوسی

memarian@kntu.ac.ir

تاریخ پذیرش: [۹۷/۳/۲۱]

تاریخ دریافت: [۹۶/۰۴/۱۶]

## چکیده:

روش‌های متداول اندازه‌گیری آلاینده‌های هوای دارای خطا، نیازمند فضای نسبتاً بزرگ و صرف هزینه‌های بسیار کلان است، در حالی که می‌توان با استفاده از روش‌های جدیدی که توانایی یادگیری دارند به آرامی از این معایب روش‌های معمول کاست. این روش‌ها که پایه ریاضی دارند و با استفاده از برنامه نویسی بنیان شده‌اند، هنوز به آن مرحله نرسیده‌اند که بتوان با اطمینان کامل جایگزین اندازه‌گیری‌های ماشینی شوند. در این مقاله از روش‌های شبکه عصبی مصنوعی و ماشین بردار پشتیبان که در پژوهش‌های گذشته نتایج قابل قبولی را برای موضوعات دیگر ارائه داده‌اند، استفاده شده است تا مقدار اوزن موجود در هوای سطح شهر تهران را با توجه به هشت پارامتر دیگر هواشناسی و آلاینده‌های معیار هوا، پیشبینی کند. در آخر با مقایسه عملکرد این دو روش با استفاده از دو معیار ارزیابی نتایج، نشان داده می‌شود که مقادیر R یا ضریب همبستگی و RMSE یا ریشه میانگین مجموع خطاها، برای ماشین بردار پشتیبان برابر است با ۰/۸۴۵۶ و ۰/۰۷۷۴ و برای شبکه عصبی مصنوعی ۰/۸۳۹۶ و ۰/۰۹۱۴، که این نتایج حاکی از برتری روش ماشین بردار پشتیبان نسبت به شبکه عصبی است. البته هر دو روش برای این پیشبینی نتایج کاملاً مطلوب و رضایت بخشی ارائه داده‌اند. همچنین میزان تاثیرگذاری پارامترها بر روی ازن تحلیل و مشخص شد که کربن مونوکسید، دمای هوا و نیتروژن دی اکسید بیشترین تاثیر را روی تغییرات ازن داشتند درحالی‌که ذرات معلق هوا و به ویژه ذرات معلق با اندازه کمتر از ۲/۵ میکرومتر کمترین تاثیر را در پیشبینی ازن داشتند.

**واژگان کلیدی:** آلودگی هوا؛ ازن تروپوسفری؛ محاسبات نرم؛ پیشبینی آلاینده

## ۱- مقدمه

کشورها در زمینه آلودگی معرفی می‌شود. شدت آلودگی هوا در شهرهای دنیا موجب شده است تا مسئولین به بررسی و مطالعه، به عنوان اولیه‌ترین راهکار مبارزه با این پدیده نگاه ویژه‌ای داشته باشند که در نتیجه در عصر کنونی در بیشتر شهرها در کشورهای در حال پیشرفت میزان آلاینده‌های هوا توسط دستگاه‌هایی در حال سنجش روزانه و حتی ساعتی است [2].

طبق گزارش سازمان جهانی سلامت در سال ۲۰۱۲ در طی یک پژوهش گسترده جهانی سالانه ۳ میلیون نفر در اثر آلودگی هوای محیط بیرون جان را خود را از دست می‌دهند که این رقم نشان دهنده خطر بسیار زیاد آلاینده‌های هوا بر روی انسان هاست [1]. در این گزارش کشور ایران به عنوان یکی از بدترین

این اندازه‌گیری‌ها نیازمند تجهیزات گسترده و با هزینه سنگین مالی است [3] و با ابداع روش‌های قابل اعتماد می‌توان از معایب اندازه‌گیری متداول فعلی کاست و میزان کارایی آن‌ها را سنجید و در مواردی می‌شود برای کالیبره<sup>۱</sup> کردن دستگاه‌ها نیز مورد استفاده قرار گیرند [4]. ازن<sup>۲</sup> یکی از مواد اصلی تشکیل دهنده لایه تروپوسفر<sup>۳</sup> یا سطحی زمین است که به این آلاینده، ازن بد یا مضر می‌گویند [5]. نوع دیگری از ازن نیز در استراتوسفر<sup>۴</sup> موجود است که به لایه ازن مشهور است و وجود الزامی برای جلوگیری از ورود اشعه‌های خطرناک خورشید به زمین دارد [6, 7]. برای اندازه‌گیری لایه ازن از تسهیلات ماهواره‌ای استفاده می‌شود ولی برای اندازه‌گیری ازن سطح زمین از دستگاه‌های اندازه‌گیری خاص ازن که دارای هزینه سنگین نیز است، بهره‌برداری می‌شود. برای اولین بار در ایران در سال ۱۳۵۳ شمسی توسط موسسه ژئوفیزیک دستگاه‌هایی برای اندازه‌گیری ازن راه اندازی شد [8]. در این پژوهش فقط به بررسی و آنالیز ازن در سطح زمین پرداخته می‌شود. تجمع میزان ازن بد در سطح زمین، به ویژه در سطح محیط‌های شهری سبب ایجاد مشکلات تنفسی از قبیل کاهش قدرت تنفس ریه، سرفه‌های متمادی، تشدید آسم و بسترسازی برای بروز بیماری‌های سخت تنفسی نظیر آسم و سرطان ریه است [9]. در یک پژوهش جامع در ایالات متحده نشان داد که ۹۵ شهر از شهرهای بزرگ این کشور، از میزان بالای ازن رنج می‌برند که این پژوهش پیشبینی کرده است که اگر یک سوم از این ازن کاسته شود می‌تواند سالانه از ۴۰۰۰ مرگ در آمریکا جلوگیری بکند [10]. همچنین طبق پژوهش در سال ۲۰۰۸ سازمان جهانی سلامت در ۲۵ کشور حوزه اتحادیه اروپا سالانه ۲۲۰۰۰ نفر بخاطر آلودگی شدید به ازن جان می‌سپارند [11]. این آمار میزان سختی مواجهه با این آلاینده و لزوم مطالعات دقیق در زمینه پیشبینی و مقابله با به خوبی نشان می‌دهد. از این رو در این پژوهش با توجه به اهمیت بالای ازن در میزان مرگ و میر ناشی از آلودگی هوا، تلاش شده است که میزان ازن هوا در سطح زمین، با بررسی‌های دقیق توسط روش‌هایی که بر مبنای یادگیری سیستم هستند،

پیشبینی شود. برخی از مطالعات پیشین مرتبط با این پژوهش بدین ترتیب است که ماشین بردار پشتیبان برای پیشبینی مقدار ازن در شهر هنگ کنگ توسط وانگ<sup>۵</sup> و لو<sup>۶</sup> برای پیشبینی تعداد روزهای پاک مورد استفاده قرار گرفته است [12]. پروفسور شلینک<sup>۷</sup> و همکاران به مقایسه ریزبینانه ۱۵ روش مختلف پیشبینی ازن پرداخته‌اند. ۸ پارامتر موثر برای این پیشبینی از کشورهای آلمان، ایتالیا، انگلیس و جمهوری چک شده‌اند. این ۱۵ روش با ۶ معیار ارزیابی مقایسه شدند که نویسندگان روش‌های شبکه عصبی و تعمیم یافته افزایشی را دارای بهترین نتیجه، در قیاس با بقیه راهکارهای پیشبینی، اعلام کرده‌اند [13]. ماشین بردار پشتیبان توسط کیشی<sup>۸</sup> و همکاران با الگوریتم کرم شب تاب<sup>۹</sup> ترکیب شده تا میزان سطح دریاچه ارومیه در افق‌های مختلف را پیشبینی کند که این روش ترکیبی در مقایسه با شبکه عصبی مصنوعی و الگوریتم ژنتیک برتری داشته است و این روش را به عنوان راهکاری قابل اطمینان برای مدیریت دریاچه ارائه داده‌اند. یگانه و همکاران از داده‌های آلاینده‌های هوا برای منطقه شهر ری در استان تهران استفاده کرده‌اند تا نشان دهند که روش ترکیبی مورد استفاده این پژوهش یعنی، ماشین بردار پشتیبان با روش حداقل مربعات جزئی<sup>۱۰</sup>، برای پیشبینی کربن مونوکسید نتایج به مراتب بهتری را نسبت به حالتی که فقط از ماشین بردار پشتیبان استفاده شده است، دارد [14]. آروین با استفاده از روش‌های آماری همبستگی و رگرسیون خطی چند متغیره، مقدار تاثیر عوامل جوی بر روی ازن شهر اصفهان را سنجید و نشان داد که تابش آفتاب و دمای هوا رابطه مستقیمی با مقدار ازن دارد و عوامل جوی نسبت به عوامل شیمیایی تاثیر بیشتری را دارند [15].

## ۲- مواد و روش‌ها

### ۲-۱- منطقه مورد مطالعه

منطقه مورد مطالعه در این پژوهش شهر تهران در مرکز کشور است و ۲۹ مین کلانشهر بزرگ دنیا با اقتصادی مشتمل بر

5 Wang

6 Lu

7 Schlink

8 Kisi

9 Firefly algorithm

10 Partial Least Square

1 Caliber

2 Ozone (O3)

3 Tropospheric

4 Stratosphere

## ۲-۲- داده‌های مورد استفاده

ایستگاه‌های سنجش کیفیت هوا و شرکت کنترل کیفیت هوای تهران در سال ۱۳۷۲ توسط شهرداری تهران تأسیس شد. این شرکت علاوه بر اجرای پروژه‌های تخصصی، مطالعاتی و پژوهشی، ۴۳ ایستگاه اندازه‌گیری صوت و ۲۲ ایستگاه سنجش آلودگی هوا در نقاط مختلف شهر تهران دارد و با بهره‌گیری از ایستگاه‌های سنجش آلودگی هوا، پارامترهای کربن مونوکسید، نیتروژن دی اکسید، سولفور دی اکسید، ازن و ذرات معلق دارای اندازه کمتر از ۲/۵ و ۱۰ میکرومتر موجود در هوای شهر تهران را اندازه‌گیری می‌کند [18]. داده‌های مورد استفاده در این مقاله از شرکت کنترل کیفیت هوا و همچنین از اداره هواشناسی کل کشور تهیه شده است. داده‌ها در این پژوهش از مناطق و محل‌های اندازه‌گیری اخذ و سپس برای هر آلاینده، مقدار بیشینه آن به عنوان مقدار آن آلاینده در آن روز انتخاب شده است چراکه در این پژوهش بدترین حالت‌های آلودگی در نظر گرفته شده است. برای داده‌های هواشناسی این مقدار میانگین اندازه‌های روزانه برداشت شده از ایستگاه‌های گوناگون در نظر گرفته شده است. داده‌های این پژوهش برای بازه زمانی از اولین روز سال ۱۳۹۳ شمسی تا آخرین روز سال ۱۳۹۴، به تعداد ۷۳۰ روز است. یعنی از هر پارامتر در این پژوهش ۷۳۰ داده موجود است که از تمامی ۲۲ ایستگاه اندازه‌گیری آلودگی هوا استفاده شده است. در این پژوهش سعی شده است از تمامی داده‌های دقیقی که می‌توانند در مقدار ازن تأثیری داشته باشند استفاده شود. هرچند برخی داده‌ها مانند ساعات آفتابی دارای تعداد زیادی داده گم شده بودند که از آنها صرف نظر شد.

## ۲-۳- پیش پردازش داده‌ها

داده‌های این پژوهش دارای واحدهای جداگانه است که برای اندازه‌گیری و مقایسه آنها با یکدیگر بدون اصلاح در داده‌ها به یک سری پیش پردازش‌ها بر روی داده‌ها نیازمندیم. هدف از پیش پردازش یا نرمال کردن، هم‌ارزش کردن تمامی عناصر موجود در یک الگو است. معمولاً نرمال کردن الگوها تبدیلی است که آنها را به بازه مشخصی مانند [۰-۱] نگاشت می‌دهد.

تجارت و صنعت دارای تعداد بسیار زیادی کارگاه و کارخانه‌های صنعتی و تولیدی است که می‌شود گفت تقریباً از هیچ کدام از حداقل‌های کنترل آلودگی هوا بهره نمی‌برند [16].

وسایط نقلیه عمومی مورد استفاده در این شهر عبارتند از: مترو، که برای تعداد نفرات استفاده‌کننده و همچنین مساحت تحت پوشش بسیار ضعیف و کم ظرفیت است و اتوبوس‌های خط واحد که آنها نیز از منابع مهم آلودگی هوا محسوب می‌شوند. تاکسی‌های شهری نیز از دیگر وسایط نقلیه مورد استفاده است. البته مردم این شهر عموماً استفاده از وسیله نقلیه شخصی را ترجیح می‌دهند و این مسئله سبب بروز ترافیک بسیار حجیمی در ساعاتی از روز می‌شود. این شهر دارای ۱۲۷۴ کیلومترمربع مساحت است که در ۲۲ منطقه تقسیم شده است که هر منطقه شهردار و مسئولین شهری مخصوص خود را دارد. تراکم جمعیت بسیار بالایی در این شهر موجود است بطوری که به ازای هر کیلومتر مربع در این شهر تقریباً ۱۳۰۰۰ نفر زندگی می‌کنند.

شکل ۱. مناطق ۲۲ گانه تهران و مکان‌های برداشت داده



Fig. 1. Data collection zones in Tehran province

مرکز شهر تهران دارای طول جغرافیایی برابر ۵۱ درجه شرقی و عرض جغرافیایی برابر با ۳۵ درجه شمالی دارد که ارتفاع از سطح دریا در این شهر از ۹۰۰ تا ۱۸۳۰ متر متغیر است. شیب این تغییر ارتفاع از شمال به جنوب است [17].

باشند یک شبکه عصبی مصنوعی معمولاً با این سه پارامتر تعریف می‌شود: ۱- الگوی اتصال بین نرون‌ها در لایه‌های مختلف، ۲- وزن ارتباطات، که در بخش یادگیری به روز می‌شوند و ۳- تابع عملگری که در نهایت نرون‌های وزن داده شده را به داده‌های خروجی تبدیل می‌کند [22]. در این روش برای یادگیری از داده‌های آموزشی و برای مقایسه داده‌های تولید شده با داده‌های مشاهداتی از داده‌های استفاده شده است. نوع شبکه انتخاب شده شبکه پیشخور با الگوریتم آموزشی پس انتشار خطا<sup>۱</sup> است که در قیاس با سایر انواع این نوع شبکه با توجه به مطالعات پیشین بسیار عملکرد بهتری دارد [23]، انتخاب می‌شود. برای نتایج بهتر باید تابعی مناسب مسئله مورد نظر انتخاب شود که در اینجا تابع TrainLM با توجه به سرعت و دقت عملکردش انتخاب می‌شود و برای تابع آموزشی تطبیقی نیز بر همان اساس تابع LEARNNGDM انتخاب شده است. تعداد لایه‌هایی مخفی یا میانی در این شبکه برابر با ۱ در نظر گرفته می‌شود و یک لایه خروجی و یک لایه ورودی در شبکه موجود است که در شکل (۲) قابل مشاهده است.

در قدم بعدی بایستی تعداد نرون‌های هر لایه را تشخیص داد که نیازمند بررسی و سعی و خطای بسیار زیاد است. البته

شکل ۲: لایه‌های مختلف برای نرون‌های شبکه عصبی

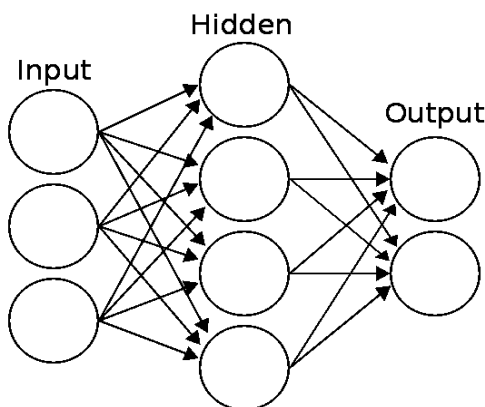


Fig. 2. Layers of neurons for ANN

تعداد نرون‌های لایه خروجی با توجه به تعداد پارامترها بصورت خودکار تنظیم می‌شود. برای تعداد نرون‌های لایه میانی از عدد ۱۰ استفاده شده است. با این حال چون فقط یک خروجی مورد نظر است برای لایه خروجی نیز بالطبع

در این پژوهش روش استفاده شده برای پیش‌پردازش داده‌ها از معادله (۱) استفاده شده است [19]:

$$X = 0.1 + 0.9 \frac{(X_i - X_{min})}{(X_{max} - X_{min})} \quad (1)$$

که در این رابطه  $X_{min}$  و  $X_{max}$  به ترتیب کمترین و بیشترین اندازه از هر پارامتر اندازه گیری شده را نشان می‌دهد و  $X_i$  مقدار داده‌ای است که می‌خواهیم نرمال سازی یا پیش پردازش را روی آن انجام دهیم.

## ۲-۴- شبکه عصبی مصنوعی<sup>۱</sup>

مغز انسان پیچیده‌ترین سیستمی است که تاکنون بشر با آن مواجه شده است. این سیستم بسیار درهم تنیده و پیچیده فقط در ۱۴۰۰ گرم مغز انسان جای گرفته است. همه‌ی فرآیندهای خودآگاه و ناخودآگاهی که در انسان رخ می‌دهد و همه دستورات حرکتی اختیاری و غیر اختیاری در مغز طراحی می‌شود. همه این قدرت مغز به ارتباطات بسیار قوی موجود در مغز برمی‌گردد، چرا که اجزای مغز خود به تنهایی از لحاظ سرعت عملکرد حدود یک میلیون بار کندتر از ترانزیستورهای مورد استفاده در CPU است [20]. برای شبیه سازی یک سیستم مشابه آنچه در مغز انسان می‌گذرد، از مدت‌ها پیش، که در کامپیوترها الگوریتم‌های محاسباتی را می‌شد پیاده‌سازی کرد، برنامه نویسان و متخصصین این امر کارهای پژوهشی بسیاری را شروع کرده‌اند، که در نتیجه این کارها، در شاخه‌ای از علم هوش مصنوعی در زیر شاخه هوش محاسباتی، تحت عنوان موضوع شبکه‌های عصبی مصنوعی طبقه‌بندی شده است. واحدهایی مصنوعی که با تقلید مغز به انتقال داده از نقاط اولیه (لایه ورودی) به نقاط ثانویه (لایه خروجی) را انجام می‌دهند، به نرون‌های مصنوعی<sup>۲</sup> معروف هستند [21]. این ارتباط بین نرون‌ها سیگنال‌هایی را دریافت و به نرون دیگر در لایه دیگر حمل می‌کنند که البته اگر سیگنال قوی‌تر باشد این انتقال بیشتر و بیشتر انجام می‌گیرد. چنین ساختارهایی می‌توانند با یادگیری، عملکرد بهتری نسبت به سایر روش‌های محاسباتی داشته

1 Artificial Neural Network (ANN)

2 Artificial Neurons

3 Feed-forward backprop

می‌شود، بی‌نهایت خط وجود دارد که این نوع داده را از هم جدا کند ولی ما به دنبال ابرخطی هستیم که بیشترین فاصله یا حاشیه را از هر دو نوع داده داشته باشد. برای داشتن بهترین خط، ابتدا هرکدام از خطوط حاشیه باید رسم شوند و سپس ابرخط را که از لحاظ هندسی دارای فرمولی  $\frac{2}{\|w\|}$  است را می‌توان رسم کرد [25]. برای شکل (۳) خطوط مرزی با توابع زیر را داریم:

$$\vec{w} \cdot \vec{x} - b = 1 \quad (2)$$

و

$$\vec{w} \cdot \vec{x} - b = -1 \quad (3)$$

که بردار  $\vec{w}$  بردار نرمال ابرخط است.

شکل ۳: خطوط مرزی و ابر برای دسته بندی داده‌ها توسط SVM

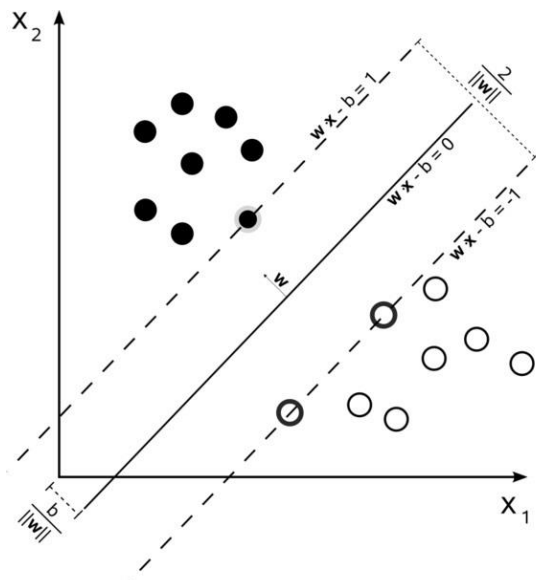


Fig. 3: Classification border lines and super line in SVM

اما در حالت‌هایی که نتوان از خط برای دسته بندی استفاده کرد، از تابعی به نام تابع کرنل<sup>۴</sup> برای نگاشت داده‌ها از صفحه (محیط دوبعدی) به فضای  $n$  بعدی استفاده می‌شود که چهار نوع تابع کرنل برای این نگاشت وجود دارد که عبارتند از: چند جمله‌ای همگن، چند جمله‌ای غیرهمگن، تابع پایه شعاعی گاوسی و تانژانت هایپربولیک. در این مقاله از روش پایه شعاعی گاوسی استفاده شده است که نتایج بهتری را در مطالعات پیشین داشته است [24].

یک نرون به صورت خودکار تولید شده است. برای توابع انتقال در لایه خروجی و در لایه میانی به ترتیب تابع خطی و تابع تانژانت سیگموئید انتخاب شده‌اند. بعد از اعمال شرایط فوق در شبکه‌ی عصبی مصنوعی، معماری و ساختار آن بارها مورد تغییر قرار داده شد و بهترین حالت‌های تجربه شده برای ساختار شبکه در نظر گرفته شد. سیستم به گردش داده‌ها در بین نرون‌ها و میان لایه‌ها و وزن دهی داده‌ها می‌پردازد و در انتها به داده‌هایی که خطای جذر مربعات را کمتر می‌کند وزن بیشتری تخصیص داده و با رسیدن به بهترین حالت ممکن، مدل‌سازی متوقف می‌شود. در این مقاله برای مدل‌سازی از طریق شبکه مصنوعی ۷۰٪ داده‌های موجود برای آموزش و ۳۰٪ باقی برای آزمون مورد استفاده قرار داده شده است. این تفکیک داده‌ها، بعد از انجام مدل‌سازی و سعی و خطا به عنوان بهترین ترکیب داده مورد استفاده قرار گرفته است.

## ۲-۵- ماشین بردار پشتیبان<sup>۱</sup>

نسخه اولیه ماشین بردار پشتیبان توسط وپنیک<sup>۲</sup> در سال ۱۹۶۳ نوشته شد که برای دسته بندی و آنالیز اعداد در روی یک صفحه بود که توسط خطوط دسته بندی انجام می‌گرفت اما خود او و کورتس<sup>۳</sup> در سال ۱۹۹۵ به تشریح پیشبرد این روش در فضاهای چند بعدی، با اعمال روش کرنل که توسط دسته بندی‌های غیرخطی انجام می‌گرفت، پرداختند [24]. دسته‌بندی داده‌ها یکی از وظایف مشترک برنامه‌های آموزشی است. به عبارت دیگر در برخی از این برنامه‌ها ابتدا دسته‌ای از داده‌ها برای آموزش ماشین و برای تولید داده‌های شبیه سازی شده استفاده می‌شود، در حالی که در دسته‌ای دیگر داده‌ها نیز برای آزمون داده‌های شبیه سازی شده استفاده می‌شوند. وقتی که داده‌ها از یک یا چند نوع هستند با استفاده از این روش به دسته‌بندی داده‌ها برحسب نوع آنها می‌پردازیم. وقتی که داده‌ها مانند شکل (۳) در یک صفحه باشند، طبقه بندی خطی خواهد بود و همان‌گونه که در شکل (۳) نیز دیده

1 Support Vector Machine

2 Vladimir Vapnik

3 Corinna Cortes

روش ماشین بردار پشتیبان و شبکه عصبی مصنوعی میزان توانایی این دو روش را مقایسه کنیم. همچنین داده‌های هر پارامتر برای آنالیز حساسیت آزن به هر پارامتر طبق بخش ۳-۳ به سیستم‌های مورد بحث وارد می‌شوند تا بهترین مدل برای هر روش انتخاب شود و میزان اثرگذاری تغییر هر پارامتر بر روی آزن سنجیده شود.

### ۳-۱- نتایج شبکه عصبی مصنوعی

در مقایسه نتایج حاصل از مدل ساخته شده توسط شبکه عصبی مصنوعی با داده‌هایی که به عنوان داده‌های آزمون به سیستم معرفی شده‌اند، شبکه عصبی طراحی شده بعد از ۵ بار تکرار توانسته است داده‌هایی را که بیشترین شباهت به داده‌های آزمون دارد را تولید کند. مقدار میانگین خطای مربعات برای بهترین مدل ساخته شده با توجه به ورودی‌ها،  $0.008376$  است که با گرفتن ریشه این عدد  $RMSE = 0.0914$  نتیجه می‌شود.

شکل ۴. خطوط رگرسیون برای نتایج شبکه عصبی مصنوعی

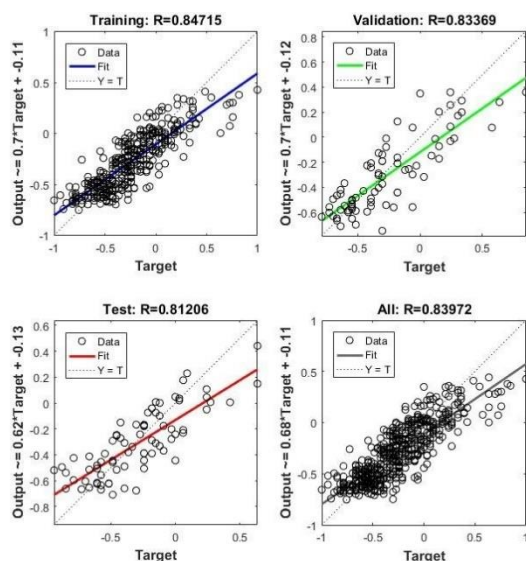


Fig. 4. Regression lines for ANN results

میزان همبستگی داده‌های تولید شده با داده‌های ورودی به شبکه عصبی در شکل (۴) قابل مشاهده است. که برای مقایسه با روش ماشین بردار پشتیبان باید مقدار همبستگی

این روش برای پیش‌بینی مقدار آزن با استفاده از ۸ پارامتر دیگر است که نشان می‌دهد دسته بندی این داده‌ها با توجه به مقدار بسیار زیاد داده‌ها و همچنین پارامترهای موثر دارای پیچیدگی بسیار زیاد است. تعداد ۲۸۶ خط بردار برای تقسیم بندی ترسیم شده است که ۹۷ عدد آن بردارهای مرزی است. در داده‌های ورودی این روش از ۳۰٪ داده‌ها به عنوان داده‌های آزمون و از ۷۰٪ آنها به عنوان داده‌های آموزشی بهره برده شده است. این تفکیک بعد از آزمون و خطاهای دیگر ترکیب‌های احتمالی استفاده شده است.

### ۲-۶- معیارهای ارزیابی و مقایسه

در این مقاله برای مقایسه هر دو روش عنوان شده در بخش ۲-۴ و ۲-۵ از معیارهای ارزیابی ضریب همبستگی<sup>۱</sup> و ریشه میانگین خطای مربعات<sup>۲</sup> برای محاسبه میزان شباهت داده‌های تولید شده و داده‌های مشاهداتی مورد استفاده قرار می‌گیرند. روابط ۴ و ۵: هرچه روشی عدد بیشتری داشته باشد نتیجه بهتری را دارد [26].

$$R = \frac{\sum_{i=1}^N (Y_m - \bar{y}_m) \times (Y_p - \bar{y}_p)}{\sqrt{\sum_{i=1}^N (Y_m - \bar{y}_m)^2} \times \sqrt{\sum_{i=1}^N (Y_p - \bar{y}_p)^2}} \quad (4)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (Y_m - Y_p)^2}{N}} \quad (5)$$

که در آن  $Y_m$  و  $Y_p$  به ترتیب میزان آزن پیش‌بینی شده و مشاهداتی و  $\bar{y}_m$  و  $\bar{y}_p$  به ترتیب میانگین آزن پیش‌بینی شده و مشاهداتی و همچنین  $N$  تعداد کل داده‌های ورودی را نشان می‌دهد.

### ۳- نتایج و بحث

طبق آنچه در بخش‌های گذشته اشاره شد، داده‌های برداشت شده مربوط به دو سال برای تخمین مقدار آزن در سطح شهر تهران با توجه به مقادیر دیگر آلاینده‌های معیار و ۳ پارامتر هواشناسی، مورد استفاده قرار گرفته‌اند تا در جایگزینی در دو

1 Correlation Coefficient

2 Root Mean Square Error

## ۳-۳- آنالیز حساسیت ازن برای تغییرات دیگر پارامترها

برای اینکه مقدار اثرگذاری تغییرات ۸ پارامتر دیگر بر روی مقدار ازن سنجیده شود، هر کدام از پارامترها در دو دسته با نام A و B وارد سیستم‌های پیشبینی شده‌اند. دسته A با هسته مرکزی سه پارامتری که با احتمال بالا روی ازن تاثیرگذارند، یعنی دمای هوا، کربن مونوکسید و نیتروژن دی اکسید و گروه B دارای هسته مرکزی ذرات معلق و رطوبت هوا هستند. در هر دو گروه، پارامترهای دیگر یکی پس از دیگری به هسته اصلی و ثابت اضافه می‌شوند که نتایج مدل‌های ساخته شده از شبکه عصبی مصنوعی در جدول (۱) و مدل‌های تولید شده توسط ماشین بردار پشتیبان در جدول (۲) قابل مشاهده است.  $A_0$  و  $B_0$  نشان دهنده هسته مرکزی برای هر دو دسته است و با هر افزایش رقم زیرنویس برای گروه A به ترتیب سولفور دی اکسید،  $PM_{10}$  و  $PM_{2.5}$ ، سرعت باد و رطوبت هوا اضافه می‌شوند. برای گروه دوم نیز با هر افزایش زیرنویس به ترتیب سرعت باد، سولفور دی اکسید، نیتروژن دی اکسید، کربن مونوکسید، و دمای هوا اضافه می‌شوند. این افزایش پارامترها به نحوی است که در ورودی‌های  $A_5$  و  $B_5$  تمامی پارامترهای برداشت شده برای پیشبینی ازن موجود باشند.

جدول ۱. مدل‌های ورودی و نتایج آن برای ANN

R	RMSE	Input	R	RMSE	Input
0.04510	0.4245	$B_0$	0.8095	0.1123	$A_0$
0.4914	0.4145	$B_1$	0.8256	0.1098	$A_1$
0.5123	0.3874	$B_2$	0.8302	0.994	$A_2$
0.5980	0.3041	$B_3$	0.8306	0.992	$A_3$
0.7074	0.1804	$B_4$	0.8355	0.0928	$A_4$
0.7778	0.1073	$B_5$	0.8396	0.0914	$A_5$

Table 1. Input models and results of ANN

جدول ۲. مدل‌های ورودی برای نتایج آن برای SVM

R	RMSE	Input	R	RMSE	Input
0.4852	0.3659	$B_0$	0.8111	0.09045	$A_0$
0.5454	0.3658	$B_1$	0.8159	0.0871	$A_1$
0.6245	0.3658	$B_2$	0.8197	0.0801	$A_2$
0.6871	0.2994	$B_3$	0.8299	0.0800	$A_3$
0.7221	0.2485	$B_4$	0.8354	0.07999	$A_4$
0.8321	0.0886	$B_5$	0.8456	0.07745	$A_5$

Table 2. Input models and results of SVM

داده‌های تولید شده با داده‌های آزمون را انتخاب کرد که برای این روش این مقدار برای بهترین سری ورودی پارامترها برابر است با  $0.081206$ . همچنین شکل (۴) نشان می‌دهد که شبکه عصبی نتیجه بهتری را در حال استفاده از داده‌های آموزشی نسبت به استفاده از داده‌های آزمون دارد و شبکه دچار آموزش بیش از حد نشده است.

در حالی شبکه دچار آموزش بیش از حد می‌شود که نتایج در قیاس با داده‌های آزمون بهتر از نتایجی باشد که در قیاس با نتایج آموزش باشد.

## ۳-۲- نتایج ماشین بردار پشتیبان

با اتمام مدل‌سازی توسط ماشین بردار پشتیبان برای کنترل آموزش بیش از حد و تصدیق کارکرد این روش باید عملکرد آن در قیاس با داده‌های آزمون و آموزشی مورد بررسی قرار گیرد. با توجه به اینکه ضریب همبستگی برای داده‌های آموزشی برابر با  $0.9320$  و برای داده‌های آزمون برابر  $0.8456$  است، مدل ساخته شده دارای این نقص نیست. خط رگرسیون ترسیم شده برای داده‌های آزمون و داده‌های مدل شده در شکل (۵) قابل مشاهده است. نتیجه بدست آمده از این مدل سازی مقدار RMSE را برای داده‌های آزمون برابر با  $0.0774$  نشان می‌دهد.

شکل ۵. خط رگرسیون برای نتایج ماشین بردار پشتیبان

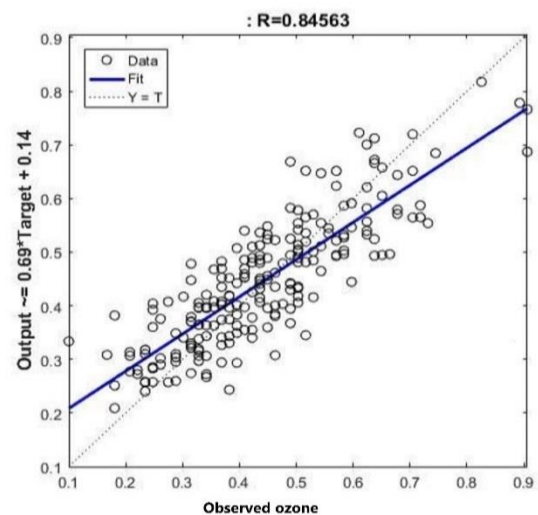


Fig. 5. Regression line of result of SVM

## 1 Over-training

## ۴- نتیجه گیری و پیشنهادات

رابطه مستقیمی دارد) و نیتروژن دی اکسید بیشترین تاثیر را در روند تغییرات آزن دارند و میزان رطوبت و سرعت، باد کمک اندک اما مثبتی برای تخمین بهتر آزن دارند. مقدار ذرات معلق دارای اندازه ۱۰ میکرون تاثیر بسیار جزئی دارند و ذرات معلق دارای اندازه ۲/۵ میکرون تقریباً در این پیشبینی تاثیری ندارند. این نتایج با پژوهشات علمی پیشین [15] همخوانی و انطباق دارد. همچنین شایان ذکر است که نور خورشید و به عبارتی دیگر ساعات آفتابی تاثیر مهمی در مقدار آزن دارد ولی با توجه به کمبود داده دقیق در بازه زمانی استفاده شده در این مقاله، پیشنهاد می شود برای پژوهشات بعدی از میزان اثرگذاری نور خورشید به عنوان پارامتر موثر در تولید آزن استفاده شود و چشم پوشی از مقدار ذرات معلق برای تسهیل روند تخمین بدون مشکل بنظر می آید. همچنین جدول ۳ نشان می دهد که ماشین بردار پشتیبان قدرت بیشتری برای شبیه سازی داده هایی مشابه به مقدار آزن اندازه گیری شده را دارد که این ناشی از توانایی ماشین بردار پشتیبان در انتقال داده ها به فضاهای چند بعدی و دسته بندی آنهاست.

جدول ۳. نتایج حاصل از بهترین مدل ها از روش های مورد مطالعه

RMSE	R	Methods/Evaluation Criteria
0.0774	0.8456	Support Vector Machine
0.0914	0.8396	Artificial Neural Network

Table 3. Results of best model of ANN and SVM

## References

## ۵- مراجع

- [1] WHO, "Health Aspects of Air Pollution with Particulate Matter, Ozone and Nitrogen Dioxide," 2003.
- [2] J. Zheng, L. Zhong, T. Wang, P. K. K. Louie, and Z. Li, "Ground-level ozone in the Pearl River Delta region: Analysis of data from a recently established regional air quality monitoring network," *Atmos. Environ.*, vol. 44, no. 6, pp. 814–823, 2010.
- [3] M. Lippmann, "Health effects of ozone a critical review," *JAPCA*, vol. 39, no. 5, pp. 672–695, 1989.
- [4] Z. Z. Liu, J. Y. Wang, and H. T. Wang, "Ozone Measurements at Some Work Places for Environment Protection and Career Health Care," *Ozone Sci. Eng.*, vol. 00, no. 00, pp. 1–6, 2017.
- [5] V. Mehdipour and M. Memarianfard,

در این پژوهش مقدار آزن موجود در سطح شهر تهران مورد مطالعه قرار گرفته است که این آلاینده به عنوان یکی از مزمن ترین مواد موجود در هوا نیازمند بررسی دقیق است. از این پژوهش می توان در کالیبره کردن دستگاه فعلی و حتی جایگزینی روش های فعلی با متودهای پیشبینی استفاده کرد. در این مقاله سعی شد از دو روش خود-یادگیر شبکه عصبی مصنوعی و ماشین بردار پشتیبان برای مدل سازی استفاده شود تا با استفاده از سری زمانی داده های برداشت شده برای کربن مونوکسید، نیتروژن دی اکسید، سولفور دی اکسید، ذرات معلق هوا با اندازه کمتر از ۲/۵ میکرومتر و ذرات معلق هوا با اندازه کمتر از ۱۰ میکرومتر، دمای هوا، رطوبت و سرعت باد به پیشبینی مقدار آزن موجود در هوا پرداخته شود. برای یکسان سازی مقادیر موجود که دارای واحدهای مختلف بودند از استاندارد سازی طبق رابطه (۱) استفاده شد که تمامی پارامترها در بازه [0-1] نگاشت شدند. در هر دو روش ۳۰٪ داده های اولیه به عنوان داده هایی که داده های تولید شده با آن ها مقایسه شوند، تخصیص داده شده اند که به اصطلاح داده های آزمون نامیده می شوند. با توجه به اینکه معیارهای ارزیابی ضریب همبستگی و ریشه میانگین خطای مربعات تعیین شده اند مقایسه این دو معیار برای روش های مذکور با توجه به جدول (۱) نشان می دهد روش ماشین بردار پشتیبان نتیجه بهتری را ارائه داده است. هرچند هر دو روش نتایج کاملاً مطلوب و قابل اطمینانی را ارائه داده اند که پیشنهاد می شود در برنامه های آتی مدیریت شهری و همچنین سایر پژوهشات مرتبط با آلاینده های هوا از این دو راه کار به عنوان روشی متمر ثمر استفاده شود. همچنین برای مقایسه و ارزیابی میزان تاثیر پارامترها بر روی مقدار آزن، با گروه بندی داده ها به دو نوع A و B و اضافه کردن یک یک دیگر پارامترها به هسته مرکزی، نشان داده شد که در هر دو روش گروه A با هسته مرکزی کربن مونوکسید، نیتروژن دی اکسید و دمای هوا بهترین نتیجه را ارائه داده است. به صورت کلی می توان از جدول ۱ و جدول ۲ نتیجه گرفت که به ترتیب کربن مونوکسید، دمای هوا (عموماً دمای هوا با ساعات آفتابی



- [20] X. Feng, Q. Li, Y. Zhu, J. Hou, L. Jin, and J. Wang, "Artificial neural networks forecasting of PM2.5 pollution using air mass trajectory based geographic model and wavelet transformation," *Atmos. Environ.*, vol. 107, pp. 118–128, 2015.
- [21] P. Martí *et al.*, "Artificial neural networks vs. Gene Expression Programming for estimating outlet dissolved oxygen in micro-irrigation sand filters fed with effluents," vol. 99, pp. 176–185, 2013.
- [22] U. Schlink *et al.*, "Statistical models to assess the health effects and to forecast ground-level ozone," *Environ. Model. Softw.*, vol. 21, no. 4, pp. 547–558, 2006.
- [23] Y. Feng, W. Zhang, D. Sun, and L. Zhang, "Ozone concentration forecast method based on genetic algorithm optimized back propagation neural networks and support vector machine data classification," *Atmos. Environ.*, vol. 45, no. 11, pp. 1979–1985, 2011.
- [24] S. Theodoridis, *Pattern Recognition*, 4th editio. Burlington: Academic Press, 2008.
- [25] C. Cortes and V. Vapnik, "Support Vector Networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.
- [26] K. Roushangar and F. Homayounfar, "Prediction of Flow Friction Coefficient using GEP and ANN Methods," *Int. J. Artif. Intell. Mechatronics*, vol. 4, no. 2, pp. 65–68, 2015.
- "Application of Support Vector Machine and Gene Expression Programming on Tropospheric ozone Prognosticating for Tehran Metropolitan," *Civ. Eng. J.*, vol. 3, no. 8, p. 557, Sep. 2017.
- [6] V. Mehdipour, "Temporal modeling of tropospheric ozone and analysis of its relationship with photochemical precursors considering meteorological parameters," K. N. Toosi University of Technology, 2017.
- [7] V. Mehdipour and M. Memarianfard, "Ground-level O3 sensitivity analysis using support vector machine with radial basis function," *Int. J. Environ. Sci. Technol.*, May 2018.
- [8] S. S. Lim *et al.*, "A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990-2010: A systematic analysis for the Global Burden of Disease Study 2010," *Lancet*, vol. 380, no. 1, pp. 2224–2260, 2012.
- [9] M. Kampa and E. Castanas, "Human health effects of air pollution," *Environ. Pollut.*, vol. 151, no. 2, pp. 362–367, 2008.
- [10] M. L. Bell, A. Mcdermott, S. L. Zeger, and J. M. Samet, "in 95 US Urban Communities, 1987-2000," *Forestry*, vol. 292, no. 19, pp. 2372–2378, 2004.
- [11] R. Gauld *et al.*, "The World Health report 2008 - Primary healthcare: How wide is the gap between its agenda and implementation in 12 high-income health systems?," 2008.
- [12] W.-Z. Lu and D. Wang, "Ground-level ozone prediction by support vector machine approach with a cost-sensitive classification scheme," *Sci. Total Environ.*, vol. 395, no. 2–3, pp. 109–116, 2008.
- [13] U. Schlink *et al.*, "A rigorous inter-comparison of ground-level ozone predictions," *Atmos. Environ.*, vol. 37, no. 23, pp. 3237–3253, 2003.
- [14] B. Yeganeh, M. S. P. Motlagh, Y. Rashidi, and H. Kamalan, "Prediction of CO concentrations based on a hybrid Partial Least Square and Support Vector Machine model," *Atmos. Environ.*, vol. 55, pp. 357–365, 2012.
- [15] A. Arvin, "Statistical Analysis of Atmosphere Parameters Effecton Surface Ozone in Isfahan city," *Urban Climatol. Res.*, vol. 1, no. 2, p. 61–70 (In Persian), 2011.
- [16] A. R. Hosseinpoor, M. H. Forouzanfar, M. Yunesian, F. Asghari, K. H. Naieni, and D. Farhood, "Air pollution and hospitalization due to angina pectoris in Tehran, Iran: A time-series study," *Environ. Res.*, vol. 99, no. 1, pp. 126–131, 2005.
- [17] "Tehran Metropolitan Atlas," *Tehran Municipality*. [Online]. Available: <http://en.tehran.ir/Default.aspx?tabid=104>.
- [18] "Tehran Air Quality Comany," *Tehran Municipality*. [Online]. Available: <http://air.tehran.ir>.
- [19] V. Mehdipour, M. Memarianfard, and F. Homayounfar, "Application of Gene Expression Programming to water dissolved oxygen concentration prediction," *Int. J. Hum. Cap. Urban Manag.*, vol. 2, no. 1, pp. 1–10, 2017.

# Temporal Prediction of Tropospheric Artificial Neural Network and Ozone Using Support Vector Machine

Vahid Mehdipour<sup>1</sup>, Mahsa Memarianfard<sup>2\*</sup>

1- Assistant professor, Faculty of Civil Engineering., K N Toosi University of Technology

2- M.Sc. Environmental Civil Engineering, K N Toosi University of Technology

memarian@kntu.ac.ir

## Abstract

Air pollution as a silent murderer of metropolitan residents, needs acute efforts. During the past few decades, after London 1954 black days, the world encountered a novel and anthropologic problem. Scientific researches for scrutinizing the air pollution sources and its effects on humankind and environment started and improved after chronic influences of contaminations have been revealed. Prognostication of pollutants and finding the relationships between parameters, seems to be undeniable. Ozone as a tropospheric gas, has severe impacts on the all creatures while the human beings are more delicate in conjunction with this gas where it can deteriorate lung ability and cause acute asthma and other pulmonary diseases. In the present article, the two most prevailing approaches for prediction, have been applied to the forecast tropospheric ozone value via considering other eight parameters. Sulfur dioxide (SO<sub>2</sub>), nitrogen dioxide (NO<sub>2</sub>), carbon monoxide (CO) and particulate matters (PM<sub>2.5</sub>, PM<sub>10</sub>) as photochemical precursors, and also relative humidity, air temperature and wind speed as meteorological parameters, after data preparation, used for ground level ozone prognostication in Tehran, Iran, with a condensed population where suffers from severe air contaminations and high rate of daily death caused air pollution. The used data series, have been collected from 22 regions of Tehran during the two years (2014 and 2015). Two evaluation criteria; root mean square error (RMSE) and correlation coefficient (R), have been selected to compare the methodologies. Support vector machine (SVM) and artificial neural networks (ANN) as capable soft computing approaches have been used in numerous areas of science, opted in this research. Support vector machine with classification of other eight parameters and by 286 vectors as a classifier and 97 border vectors, sorted the 70 percent of data sets as training and the residual amount of parameters used as testing data sets. Radial basis function (RBF) selected as the Kernel function. Artificial neural network works as like as human brain and neurons between layers transfer datasets and process them during the run time, where in the recent paper the layer number of the created network is one for hidden layer and one for the output layer and 10 neurons have been exploited for the hidden layer and one for the output layer. Network type of this system is feed-forward with back propagation and TRAINLM used as training function and LEARNGDM used for adaption learning function. Both approaches depicted reliable and acceptable results, where RMSE and R values for support vector machine, respectively are 0.0774 and 0.8456, also artificial neural network resulted 0.0914 for RMSE and 0.8396 for R, which are reasonable outcomes. As the outcomes for training datasets were better than the results for testing datasets, both approaches showed that both of they did not over-trained, which is a serious and prevalent difficulty for soft computers. Support vector machine, with lower root mean square error and higher correlation coefficient disclosed as the better method for ground level ozone prediction. These series of studies are supportive for calibration of measuring systems and due to their expensiveness, soft computing can be the most reliable and affordable substitute. Also the analysis of tolerances among the parameters illustrated that CO, Temperature and NO<sub>2</sub> are the most effective where, PM<sub>2.5</sub> had the least amount impact on O<sub>3</sub> forecasting process.

**Keywords:** Air pollution; Tropospheric ozone; soft computing; prediction; Tehran.