

## Research Paper

## A wavelet-ANN-based framework for estimating air pollutant concentrations using remotely sensed data in Tehran metropolitan area

Ali Shamsoddini<sup>\*1</sup>, Mohammadreza Aboodi<sup>2</sup>

1. Assistant Professor, Department of Remote Sensing, Tarbiat Modares University, Tehran, Iran.
2. Phd Student of Remote Sensing, Tarbiat Modares University, Tehran, Iran.

**Received:** 14 August 2020**Accepted:** 12 April 2021**PP:** 115-130

Use your device to scan and read the article online

**Keywords:**

Air pollution, Modeling, Random forest, Wavelet transform, Artificial neural network Multilayer perceptron

**Abstract**

In developing countries, most major cities are increasingly exposed to air pollution as a factor affecting the quality of life and public health of the community. High population density in Tehran causes this metropolitan area to be one of the most important region in Iran. Polluting industry and the use of polluting transportation are two of the main sources of air pollutant in Tehran and have turned this city to the most polluted metropolitan area in Iran. Consequently, the need for the air pollution reduction is too necessary in this area. The air pollutant concentration predictions can improve decision making for appropriate solutions to reduce air pollution. This study examined the performance of the random forest feature selection and wavelet transformation methods when they combine with the multiple-linear regression and multilayer perceptron artificial neural network to achieve an efficient model to estimate several pollutants including carbon monoxide, nitrogen dioxide, sulfur dioxide, and PM<sub>2.5</sub> in Tehran metropolitan area. For these purpose four groups of remotely sensed-derived and spatial data including spatial data, meteorological data, traffic information, and the air pollutant concentrations in the days before the prediction day were applied as the input data of the models. Results showed that the modeling of all pollutants by the multilayer perceptron neural network along with the wavelet transform method provides higher accuracy than the other models. Furthermore, the estimation accuracy of the carbon monoxide pollutant (with error of estimation=19.8% ) was lower than the other pollutants while PM<sub>2.5</sub> (with error of estimation=17.0%) was estimated with higher accuracy compared to that derived for other pollutants. Moreover, it was shown that the pollutant concentrations for the days before the day for that the estimation is implemented are the most important attributes, according to the random forest feature selection method.

**Citation:** Shamsoddini, A; Aboodi, M(2022): A wavelet-ANN-based framework for estimating air pollutant concentrations using remotely sensed data in Tehran metropolitan area, Journal of Regional Planning, Vol 12, No 47, PP:115-130.

**DOI:** 10.30495/JZPM.2021.25725.3717

**\*Corresponding author:** Ali Shamsoddini**Address:** Assistant Professor, Department of Remote Sensing, Tarbiat Modares University, Tehran, Iran.**Tell:** +989155034522**Email:** ali.shamsoddini@modares.ac.ir

## Extended Abstract

### Introduction

Increase in population along with urbanization, resulted in harmful consequences for cities. Continuity of this type of urban growth can be an alarm of urban instability causing socio-economic and environmental problems. Air quality is an important parameter in urban areas, especially populated ones. Air pollution is among important environmental disasters which are considered as a serious threat of human's life. In developing countries, most of the large cities counters with air pollution affecting the public health of the community. Attending to air pollution as a factor which affects citizens during a day, is necessary for urban sustainable development. It seems that it is required to study the past trend of the air quality and predict the air pollutant concentrations for the coming days for assessing the plans conducted to control and decrease the air pollution. This study aims to design an optimum model for air pollution prediction. To achieve this goal, two preprocessing methods including random forest feature selection and wavelet transformation are applied to improve predictions. In this research, the performance of artificial neural network as a non-linear model is compared with multiple-linear regression as a linear method for predicting carbon monoxide, nitrogen dioxide, sulfur dioxide and PM<sub>2.5</sub> in 5 air quality measurement stations in Tehran. Accordingly, the research question is as follows: which is preprocessing algorithm more efficient to improve the performance of the air pollution prediction?

### Methodology

In this research, four groups of parameters affecting the air pollutant concentrations were applied. These groups include traffic factor, the pollutant concentrations measured for the last days, meteorological data and spatial factors. The meteorological parameters comprise of temperature, relative humidity, wind speed, wind direction, cloud cover, and surface pressure were used as mean daily values. These data were downloaded from Ecmwf (<http://www.ecmwf.int>). The concentrations of carbon monoxide, nitrogen dioxide, sulfur dioxide, and PM<sub>2.5</sub> measured by 5 air quality

measurement stations (including Shahid Beheshti, Cheshmeh, Elm-o-Sanat, Razi Park, and Pasdaran) in 2012 were provided by the Department of Environment. Traffic index which was proposed for 1km-radius of the stations in this study, was used as traffic data. Spatial data including UTM coordinates of the stations and mean value of NDVI for 1km-radius of the stations were utilized. In this research, multilayer perceptron (MLP) artificial neural network and multiple-linear regression fed by preprocessed and non-preprocessed data, were used for modeling air pollutant concentration. Random forest feature selection and wavelet transformation were used for preprocessing of the input data. In order to model the pollutant concentrations, Meteorological, traffic, and spatial data along with mean and maximum values of pollutant concentrations measured for last days (up to 4 days) were applied as input data, whereas daily maximum pollutant concentrations were utilized as output of the model. In this study, 1780 samples measured by 5 air quality measurement stations, were used. These samples were randomly grouped into three parts including training data (70%), validation data used for optimizing MLP parameters (15%), and test data (15%). After training, the performance of the models is assessed by the test data. The suitable values for neuron number, iteration, error threshold, learning rate, and momentum were determined by applying trial and error for each air pollutant. Mean and maximum Pollutant concentration values of the last days, were decomposed by the wavelet transformation into 5 levels, before being used as input. The final models were statistically assessed by the Paired-samples t-test to determine which method statistically performs better than the others. Moreover, independent-samples t-test was used to determine whether the amount of the pollutant affects the error of the prediction. For this purpose, the predictions of test samples were divided into two groups including samples whose values are more than mean value and lower than mean value.

### Results and Discussion

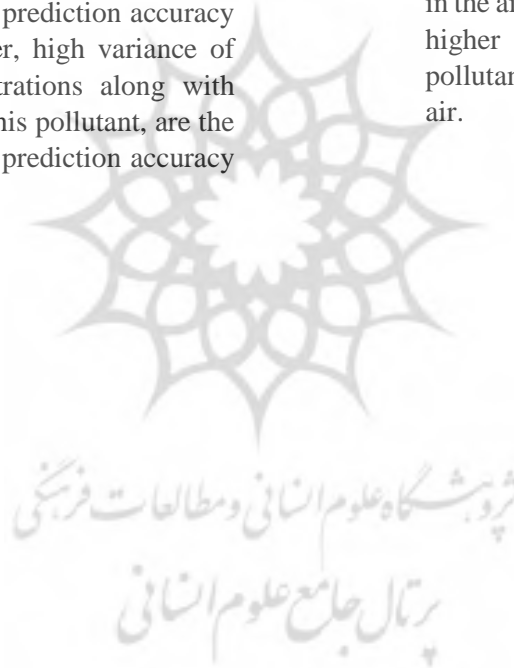
According to the results, pollutant concentrations for last days are the most important attributes for predicting the pollutant

concentration for the coming days. This issue seems to indicate that air pollutants are stable in the air and consequently, the pollutant concentrations of the last days can affect the pollutant concentrations of the next days. For carbon monoxide, nitrogen dioxide, and sulfur dioxide, independent-samples t-test showed that error prediction for the samples whose pollutant concentration values are more than mean is higher than the others due to their higher variance and consequently, complexity of these data. As the statistical tests demonstrated the prediction error of carbon monoxide is more than the other air pollutants whereas PM<sub>2.5</sub> predictions conducted by wavelet transformation-MLP, showed the lowest error of estimation (17.04%) and highest R<sup>2</sup> (0.9). Simple and linear relationship between input parameters and PM<sub>2.5</sub> concentration values can result in higher prediction accuracy of this pollutant. Moreover, high variance of carbon monoxide concentrations along with more complex sources of this pollutant, are the main reasons for its lower prediction accuracy compared to the others.

### Conclusion

This research can be briefly concluded as follows:

- Wavelet transformation/MLP models provide better accuracy for the prediction of the air pollutant concentrations.
- Carbon monoxide is predicted with the lowest accuracy and PM<sub>2.5</sub> prediction shows the lowest error among the air pollutants.
- There is no significant difference between linear and non-linear models for predicting PM<sub>2.5</sub>, sulfur dioxide, and nitrogen dioxide.
- The prediction error increases when carbon monoxide, sulfur dioxide, and nitrogen dioxide concentration increase in the air. Also, it is expected to achieve higher prediction accuracy when the pollutant concentrations are low in the air.



## ارائه یک چارچوب جهت پیش‌بینی غلظت آلاینده‌های هوا با استفاده از داده‌های سنجش‌ازدور مبتنی بر شبکه عصبی مصنوعی - موجکی در کلان‌شهر تهران

علی شمس‌الدینی<sup>۱</sup>؛ محمدرضا عبودی<sup>۲</sup>

۱. استادیار گروه سنجش‌ازدور و سیستم اطلاعات جغرافیایی، دانشگاه تربیت‌مدرس، تهران، ایران.

۲. دانشجوی دکتری سنجش‌ازدور و سیستم اطلاعات جغرافیایی، دانشگاه تربیت‌مدرس، تهران، ایران.

### چکیده

در کشورهای درحال توسعه، اکثر شهرهای بزرگ به طور فزاینده‌ای با آلودگی هوا به‌عنوان عاملی تأثیرگذار بر کیفیت زندگی و سلامت عمومی جامعه مواجه هستند. ناحیه کلان‌شهری تهران نیز به‌واسطه تراکم جمعیتی بالا یکی از مناطق مهم در ایران محسوب می‌گردد. صنایع آلاینده به همراه استفاده از وسایل نقلیه از جمله مهم‌ترین عواملی محسوب می‌گردد که موجب شده است تا این شهر به‌عنوان آلوده‌ترین کلان‌شهر کشور محسوب گردیده و متعاقباً لازم است تا در جهت کاهش آلودگی هوا در این شهر برنامه‌ریزی مؤثری انجام شود. در این پژوهش، پارامترهای تأثیرگذار بر غلظت آلاینده‌ها در قالب ۴ دسته عوامل ترافیکی، غلظت آلاینده‌ها در روزهای قبلی، داده‌های هواشناسی و عوامل مکانی به‌عنوان ورودی مدل‌ها مورد استفاده قرار گرفتند و ماکزیم غلظت آلاینده‌ها در هر روز به‌عنوان خروجی مدل در نظر گرفته شد. هدف این پژوهش بررسی عملکرد روش‌های انتخاب ویژگی جنگل تصادفی و تبدیل موجک در ترکیب با روش‌های رگرسیون چندمتغیره و شبکه عصبی مصنوعی پرسپترون چندلایه، جهت دستیابی به مدلی کارا به‌منظور پیش‌بینی آلاینده‌های منو اکسید کربن، دی‌اکسید نیتروژن، دی‌اکسید گوگرد و  $PM_{2.5}$  می‌باشد. نتایج به‌دست‌آمده نشان داد که مدل‌سازی همه آلاینده‌ها با استفاده از شبکه عصبی پرسپترون چندلایه در ترکیب با روش تبدیل موجک صحت بالاتری را نسبت به مدل‌های دیگر ارائه می‌نماید. همچنین صحت پیش‌بینی آلاینده منو اکسید کربن (خطای استاندارد برابر با ۱۹/۸ درصد) نسبت به آلاینده‌های دیگر پایین‌تر بود درحالی‌که صحت پیش‌بینی آلاینده  $PM_{2.5}$  (خطای استاندارد برابر با ۱۷/۰ درصد) بالاتر از سایر آلاینده‌ها بود. علاوه بر این، باتوجه به پارامترهای انتخاب شده توسط روش انتخاب ویژگی با استفاده از جنگل تصادفی، پارامترهای غلظت آلاینده‌ها در روزهای قبل از اهمیت بالایی به‌منظور پیش‌بینی آلاینده‌های مختلف برخوردارند.

تاریخ دریافت: ۲۴ مرداد ۱۳۹۹

تاریخ پذیرش: ۲۳ فروردین ۱۴۰۰

شماره صفحات: ۱۱۵-۱۳۰

از دستگاه خود برای اسکن و خواندن مقاله به صورت آنلاین استفاده کنید



### واژه‌های کلیدی:

مدل‌سازی، آلودگی هوا، جنگل تصادفی، تبدیل موجک، شبکه عصبی مصنوعی پرسپترون چندلایه.

**استناد:** شمس‌الدینی، علی؛ عبودی، محمدرضا (۱۴۰۱): ارائه یک چارچوب جهت پیش‌بینی غلظت آلاینده‌های هوا با استفاده از داده‌های سنجش‌ازدور مبتنی بر شبکه عصبی مصنوعی - موجکی در کلان‌شهر تهران، فصلنامه برنامه‌ریزی منطقه‌ای، سال ۱۲، شماره ۴۷، مردودشت: صص ۱۱۵-۱۳۰.

DOI: 10.30495/JZPM.2021.25725.3717

†

نویسنده مسئول: علی شمس‌الدینی

نشانی: استادیار گروه سنجش‌ازدور و سیستم اطلاعات جغرافیایی، دانشگاه تربیت‌مدرس، تهران، ایران.

تلفن: ۰۹۱۵۵۰۳۴۵۲۲

پست الکترونیکی: ali.shamsoddini@modares.ac.ir



## مقدمه

آن‌ها با استانداردها و طرح راهکارهای اجرایی کاهش آلودگی، گامی مؤثر می‌باشد (Shamsoddini et al., 2017). تحقیق حاضر باهدف طراحی مدلی بهینه جهت پیش‌بینی آلاینده‌های هوا، بر آن است تا به کمک روش‌های جنگل تصادفی و تبدیل موجک، صحت مدل‌های پیش‌بینی‌کننده آلودگی هوا را در شهر تهران بهبود بخشد. بدین ترتیب هدف اصلی این تحقیق، مقایسه عملکرد مدل توسعه‌یافته شبکه عصبی به‌عنوان روشی غیرخطی با مدل رگرسیون چندمتغیره به‌عنوان روشی خطی و بررسی امکان بهبود عملکرد مدل‌سازی آلاینده‌های منو اکسید کربن، دی‌اکسید نیتروژن، دی‌اکسید گوگرد و  $PM_{2.5}$  در ایستگاه آلاینده سنج در سطح شهر تهران از طریق استفاده از الگوریتم‌های پیش‌پردازش از جمله تبدیل موجک و جنگل تصادفی، می‌باشد. براین اساس، سؤال اصلی این پژوهش آن است که کدام‌یک از روش‌های پیش‌پردازش مورد استفاده در این تحقیق می‌تواند در کنار روش‌های مدل‌سازی، در مدل‌سازی آلاینده‌های هوای شهر تهران کارایی بیشتری داشته باشد؟

## پیشینه تحقیق و مبانی نظری

یکی از مهم‌ترین اقدامات در زمینه مدیریت و کاهش آلودگی هوا، پیش‌بینی مقادیر غلظت آلاینده‌ها می‌باشد که این امر به علت پیچیدگی فرایند و دخالت عوامل غیرقطعی در آن، مشکلاتی را به همراه دارد. مدل‌سازی آلودگی هوا برای شبیه‌سازی و وضعیت آلودگی اتمسفر با دو هدف اصلی شامل افزایش دانش در این زمینه و پیش‌بینی قابل‌اعتماد صورت می‌گیرد. اهداف ذکر شده ممکن است در برنامه‌ریزی استراتژیک و تصمیم‌گیری‌های زیست‌محیطی مورد استفاده قرار گیرد (دارای ماهیت بلندمدت) و یا در کوتاه‌مدت برای هشدار به شهروندان و مدیران استفاده شوند (Karatzas & Kaltsatos, 2007). از این‌رو در دهه‌های اخیر، توجه زیادی به مدل‌هایی شده است که قابلیت ارتباط مستقیم بین داده‌های ورودی و خروجی بدون نیاز به درک کامل علل و اثرات مربوط به فرایندهای تشکیل و پراکندگی آلاینده‌های هوا را فراهم آورند (Duro et al., 2016). از جمله روش‌های محاسباتی که در مدل‌سازی آلودگی هوا مورد استفاده قرار گرفته‌اند می‌توان به مدل شبکه عصبی پرسپترون چندلایه توسط گریواس و چالولاکو (2006)، پرز و تریر (2011)، فرناندو و همکاران (2012)، مدل شبکه عصبی توابع شعاعی پایه متوسط وانگ و همکاران (2003)،

افزایش جمعیت به همراه افزایش شهرنشینی، پیامدهای زیان‌باری برای شهرها به ارمغان آورده است. تلاوم این‌گونه رشد شهرنشینی با مشکلات اجتماعی، اقتصادی و زیست‌محیطی بحران‌آفرین توأم بوده و هشدار بر ناپایداری شهرها می‌باشد (Ahadnejad et al., 2020: 33). کیفیت هوا به‌عنوان یک عامل مؤثر بر کیفیت زندگی در مناطق شهری به‌ویژه در مناطق پرجمعیت مطرح می‌باشد. در دهه‌های اخیر تغییرات زیست‌محیطی، چالش اصلی را در ارتباط با امنیت انسانی در جهان ایجاد کرده است که بخشی از این ناامنی ناشی از تأثیر پدیده‌های اقلیمی از جمله آلودگی هوا در حوزه سلامت می‌باشد (Ahadi et al., 2019: 131). آلودگی هوا از جمله مهم‌ترین آسیب‌های زیست‌محیطی محسوب می‌گردد که تهدیدی جدی برای سلامت انسان به‌شمار می‌آید. اکثر شهرهای بزرگ کشورهای در حال توسعه، به‌طور فزاینده‌ای با آلودگی هوا به‌عنوان عاملی تأثیرگذار بر سلامت عمومی جامعه مواجه هستند. برای توسعه پایدار شهری توجه به آلودگی‌ها، به‌ویژه آلودگی هوا که شهروندان به‌صورت تمام‌وقت با آن در تماس هستند، ضروری است (Esmailnejad et al., 2015: 133). آلودگی هوای شهری متأثر از شرایط اقلیمی (موقعیت جغرافیایی، میزان دما، رطوبت هوا، وزش باد، میزان بارندگی)، افزایش جمعیت و رشد سریع شهرنشینی، منابع آلوده‌کننده، استفاده بیش‌ازحد از سوخت‌های فسیلی و ترافیک در شهرها و الگوی نامناسب سیستم‌های حمل‌ونقل می‌باشد (Arhami et al., 2013: 4777; Tavakoli et al., 2014: 75; Cai et al., 2009: 32). برطبق گزارش اتحادیه اروپا، آلودگی هوای شهری تا سال 2050، جایگاه بالاتری را در مسائل زیست‌محیطی نسبت به بحران کمبود آب و بهداشت محیط کسب خواهد کرد. (Dunea et al., 2015: 1)

داشتن تصویری روشن از آینده‌ی وضعیت آلودگی هوا بی‌تردید به برنامه‌ریزی و تصمیم‌سازی در جهت جلوگیری از وقوع بحران زیستی و یا کاهش اثرات مخرب آن کمک می‌نماید. به نظر می‌رسد مطالعه بر روی روند گذشته کیفیت هوا و نیز پیش‌بینی وضعیت آلاینده‌ها در هوا در طی روزهای بعد، می‌تواند معیاری برای ارزیابی اقدامات انجام‌گرفته در زمینه کنترل آلودگی هوا و برنامه‌ریزی برای کاهش آلودگی هوا در آینده فراهم آورد. در جهت شناسایی و رفع مشکل آلودگی هوا، عملیاتی از جمله شناسایی و پایش انواع آلاینده‌ها و منابع تولیدکننده آن‌ها، تطبیق

<sup>1</sup>Radial basis function

<sup>2</sup>Wang et al.

<sup>3</sup> Multilayer perceptron

<sup>4</sup>Grivas & Chaloulakou

<sup>5</sup>Perez and Trier

<sup>6</sup>Fernando et al.

داده‌های ورودی شبکه عصبی مربوط به غلظت  $PM_{10}$  بین سال‌های ۱۹۹۹ تا ۲۰۰۶، با استفاده از الگوریتم ژنتیک بهینه شدند. میانگین خطای مطلق پیش‌بینی شده، ۱۰ درصد به دست آمد که سه برابر بهتر از پیش‌بینی‌های حاصل از رگرسیون خطی چندمتغیره و روش رگرسیون مؤلفه‌های اصلی بود (Antanasijević et al., 2013:511). محمدی و همکاران (۲۰۱۶) در پژوهشی تحت عنوان پیش‌بینی غلظت آلاینده‌های گازی در هوای شهر تبریز، از شبکه عصبی مصنوعی به منظور پیش‌بینی غلظت آلاینده‌های  $CO$ ،  $SO_2$  و  $NO$  استفاده نمودند. بر اساس نتایج به دست آمده در این تحقیق، روش استفاده شده به عنوان روشی قابل اعتماد برای پیش‌بینی غلظت ساعتی آلاینده‌های هوا بوده که ضریب همبستگی بالای ۰٫۹ بین غلظت آلاینده‌های پیش‌بینی شده و غلظت واقعی آلاینده‌ها به دست آمد. (Mohammadi et al., 2016). نورانی و همکاران (۲۰۱۹) در مطالعه‌ای به پیش‌بینی غلظت آلاینده‌های  $SO_2$  و  $NO$  هوای شهر تبریز با استفاده از شبکه‌های عصبی مصنوعی و سیستم استنتاج عصبی فازی انطباقی پرداختند. در این مطالعه از دما، سرعت باد و رطوبت به صورت ساعتی و روزانه به همراه غلظت آلاینده‌ها استفاده گردید. نتایج این مطالعه حاکی از این بود که هر دو روش استفاده شده نتایج مناسبی را ارائه دادند (Noorani et al., 2019). در سال‌های اخیر روش‌های آماری مختلفی از جمله رگرسیون خطی چندمتغیره گام‌به‌گام (SMLR) و شبکه‌های عصبی مصنوعی به منظور پیش‌بینی آلودگی هوا مورد استفاده قرار گرفته‌اند (Wang et al., 2015:1202). مزیت روش رگرسیون چندمتغیره گام‌به‌گام سادگی اعمال این روش بوده و به همین خاطر در مطالعات مربوط به مدل‌سازی آلودگی هوا به عنوان روشی رایج جهت مدل‌سازی استفاده می‌شود (Barrero et al., 2006:67). روش‌های مختلف شبکه عصبی مصنوعی، پیچیده‌تر ولی در عین حال قابل انعطاف‌تر از مدل‌های آماری خطی هستند و قادر به توصیف داده‌هایی با وابستگی غیرخطی بالا می‌باشند. روش‌های مختلف شبکه عصبی مصنوعی اجازه می‌دهند تا از داده‌های مبهم و غیرخطی اطلاعات استخراج شوند. دومین مزیت مدل‌های شبکه‌های عصبی مصنوعی توانایی آن‌ها برای یادگیری روابط پیچیده و ناشناخته از داده‌های آموزشی است (Zhang et al., 2012:64). غلظت آلاینده‌های هوا، تحت تأثیر ترکیبی از آلاینده‌های انتشار یافته، فرایندهای فیزیکی و شیمیایی در اتمسفر و خصوصیات سطح زمین می‌باشد (Sharma et al., 2002:374).

مدل رگرسیون خطی چندمتغیره<sup>۱</sup> توسط باررو و همکاران (۲۰۰۶)، استادلور و همکاران<sup>۲</sup> (۲۰۰۸)، ماشین بردار پشتیبان<sup>۳</sup> توسط او سو سکی و همکاران<sup>۴</sup> (۲۰۰۷)، ارائه شده است، اشاره نمود. اگرچه در زمینه مدل‌سازی آلودگی هوا، با روش‌های خطی و غیرخطی مطالعات زیادی صورت گرفته است، اما به دلیل پیچیدگی ساختار آلاینده‌های هوا و رسیدن به مدل‌های دقیق‌تر و انتخاب پارامترهای مناسب برای پیش‌بینی آلاینده‌ها، مطالعات بیشتری باید انجام پذیرد.

در سال‌های اخیر در زمینه پیش‌بینی زمانی - مکانی غلظت آلاینده‌های موجود در هوا مطالعات گوناگونی صورت گرفته است. صدر موسوی و رحیمی (۲۰۱۰) در پژوهشی به پیش‌بینی غلظت ازن در شهر تبریز با استفاده از روش‌های شبکه عصبی پرسپترون چندلایه و رگرسیون خطی چندمتغیره پرداختند. بدین منظور از داده‌های ازن در ماه‌های آذر و دی سال ۱۳۸۵ به صورت ساعتی استفاده شد. نتایج به دست آمده نشان داد که مدل شبکه‌های عصبی پرسپترون چندلایه عملکرد بهتری نسبت به روش رگرسیون خطی چندمتغیره داشته است. با توجه به نتایج حاصل از این تحقیق، ضریب همبستگی به دست آمده از پیش‌بینی مدل رگرسیون چندمتغیره، ۰/۴۵ بود در حالی که ضریب همبستگی این پیش‌بینی با استفاده از مدل شبکه‌های عصبی ۰/۹۱ بود (Sadr Mousavi & Rahimi, 2010:65). گریواس و چالولاکو (۲۰۰۶) در مطالعه‌ای به ارزیابی توانایی شبکه عصبی برای ارائه پیش‌بینی‌های قابل اعتماد از عنصر  $PM_{10}$  پرداختند. در این پژوهش از داده‌های غلظت ساعتی  $PM_{10}$  در طی سال‌های ۲۰۰۱ تا ۲۰۰۲ در ۴ ایستگاه در شهر آتن یونان استفاده شد و مدل‌های شبکه عصبی مصنوعی با استفاده از ترکیب متغیرهای هواشناسی و در مقیاس زمان به عنوان ورودی، توسعه داده شدند. همچنین یک روش بهینه‌سازی بر مبنای الگوریتم ژنتیک برای انتخاب متغیرهای ورودی مورد بررسی قرار گرفت و متغیرهای دما، رطوبت نسبی، سرعت و جهت باد انتخاب شدند. نتایج حاصل از شبکه عصبی پرسپترون چندلایه که در آن به عنوان ورودی از تمام داده‌ها استفاده گردید، در تمام ایستگاه‌ها نسبتاً رضایت‌بخش بود و شاخص توافق ژین ۰/۸ تا ۰/۸۹ به دست آمد. همچنین عملکرد مدل‌های شبکه عصبی در مقایسه با رگرسیون خطی چندمتغیره برتری محسوس داشت (Grivas & Chaloulakou, 2006:1216).

آنتاناسیجیویچ و همکاران (۲۰۱۳) در مطالعه‌ای به توسعه مدل شبکه عصبی مصنوعی برای پیش‌بینی  $PM_{10}$  پرداختند.

<sup>۱</sup>Osowski & Iougaranty

Index of agreement

<sup>۳</sup>Stepwise multiple-linear regression

<sup>۱</sup>Multiple linear regression

<sup>۲</sup>Stadluber

<sup>۳</sup>Support vector machine

فشار سطح به‌صورت میانگین روزانه برای منطقه مورد مطالعه می‌باشد که از سایت Ecmwf (<http://www.ecmwf.int>) استخراج شدند. در ضمن شاخص‌های سرعت و جهت باد از طریق محاسبه مؤلفه‌های رابطه ۱ با هم ادغام گردیدند (Siwek & Osowaski., ۲۰۱۲:۱۲۴۶):

$$W_x = |W| * \cos \varphi \quad \text{رابطه (۱)}$$

$$W_y = |W| * \sin \varphi$$

در رابطه ۱،  $W$  معرف سرعت باد و  $\varphi$  معرف جهت باد می‌باشد. به‌منظور بهبود عملکرد مدل‌ها، پارامتر ماه در سال برای ارائه تغییرات آلاینده‌ها به‌واسطه تغییرات شرایط اتمسفر در هر ماه سال، با استفاده از رابطه ۲ محاسبه گردید (Arhami et al., 2013:4777):

$$MOY = \cos\left(\frac{2\pi m}{12}\right) \quad \text{رابطه (۲)}$$

در رابطه ۲،  $m$  بیانگر عدد ماه مورد نظر می‌باشد. غلظت آلاینده‌های منوکسید کربن، دی‌اکسید نیتروژن، دی‌اکسید گوگرد و  $PM_{2.5}$  ایستگاه‌های واقع در شهر تهران برای سال ۱۳۹۱ شامل ۵ ایستگاه ثابت پایش آلودگی هوا (ایستگاه‌های شهید بهشتی، چشمه، علم و صنعت، پارک رازی و پاسداران) از طریق سازمان حفاظت محیط‌زیست تهیه گردید. داده‌های ترافیکی شامل شاخص ترافیکی تا شعاع ۱ کیلومتری در هر ایستگاه بود که در این مطالعه تعریف گردید. واضح است که غلظت آلاینده‌ها با فاصله از جاده تغییر می‌کند؛ بنابراین برای به‌دست‌آوردن عدد شاخص ترافیکی هر بافر، فاصله خیابان‌ها از هر ایستگاه محاسبه شده و در عرض خیابان ضرب شد و در نهایت میانگین وزنی آن‌ها محاسبه گردید. داده‌های مکانی استفاده شده در این تحقیق شامل مختصات ایستگاه‌ها، و میانگین پوشش گیاهی تا شعاع ۱ کیلومتری از هر ایستگاه می‌باشند. به‌منظور محاسبه پارامتر میانگین پوشش گیاهی، با استفاده از تصاویر ماهواره لندست ۷ در منطقه مورد مطالعه، ابتدا از هر ایستگاه به فواصل ۱۰۰ متر تا شعاع ۱ کیلومتر بافرهایی ترسیم شد و میانگین پوشش گیاهی در هر شعاع محاسبه گردید. در این تحقیق از روش‌های شبکه عصبی پرسپترون چندلایه (MLP) و رگرسیون خطی چندمتغیره به‌منظور مدل‌سازی آلاینده‌ها به‌صورت مجزا و همچنین در تلفیق با روش‌های تبدیل موجک<sup>۱</sup> و انتخاب خصیصه جنگل تصادفی<sup>۲</sup> استفاده گردیده است. جهت انجام مدل‌سازی، پارامترهای معرفی شده در

غلظت و ترکیب آلاینده‌های هوا توسط عواملی همچون شرایط آب‌وهوایی، تأثیرات توپوگرافی و منابع انتشار، تحت تأثیر قرار می‌گیرند (Moustris et al., 2013:1634). عوامل هواشناسی، به‌طور کلی نقش مهمی در روند گسترش آلاینده، انتقال و واکنش شیمیایی آن‌ها ایفا می‌کنند و به صورت مستقیم و غیرمستقیم کیفیت هوا را تحت تأثیر قرار می‌دهند. (Feng et al., 2015:118). عوامل هواشناسی ارتباط نزدیکی با منشأ تولید، انتقال و انتشار آلاینده‌ها و همچنین تشکیل شرایط مطلوب برای تجمع و پراکنده‌سازی آلاینده‌ها دارند (Lee et al., 2011:7240). با توجه به تنوع عوامل تأثیرگذار بر میزان غلظت آلاینده‌های هوا، نیاز به روش‌های دقیق‌تر برای پیش‌بینی آلاینده‌های هوا جهت مدیریت بهتر آن وجود دارد، به نظر می‌رسد استفاده از روش‌های ترکیبی جهت مدل‌سازی آلاینده‌ها می‌تواند اقدامی مهم در این راستا باشد. با توجه به ماهیت پیچیده و غیرخطی داده‌های آلاینده‌های هوا و نیز عوامل متعدد تأثیرگذار بر آنها، از روش شبکه عصبی مصنوعی به‌منظور پیش‌بینی این آلاینده‌ها استفاده می‌شود. همچنین با توجه به ابعاد وسیع داده‌ها و پیچیدگی ساختاری آن‌ها، لازم است تا از روش‌های کاهش ابعاد ویژگی به‌منظور کاهش پیچیدگی داده‌ها استفاده شود. همچنین، به دلیل تغییرات زیاد داده‌های غلظت آلاینده‌ها، پیش‌بینی آن‌ها می‌تواند بسیار مشکل باشد و یک راه حل برای این معضل، تجزیه سیگنال‌های داده‌ها در سطوح مختلف می‌باشد (Siwek & Osowaski, 2012:1246). به نظر می‌رسد استفاده از روش‌های ترکیبی که در آن‌ها پیش از انجام مدل‌سازی از روش‌های جنگل تصادفی به‌عنوان روشی کارا در زمینه کاهش ابعاد داده‌های ورودی (Shamsoddini et al., 2014) و تبدیل موجک به‌عنوان یک روش تجزیه داده‌های ورودی به‌منظور ایجاد ویژگی‌هایی کارا (Siwek & Osowaski, 2012:1246) استفاده می‌کنند، می‌تواند به پیش‌بینی بهتر آلاینده‌های هوا کمک نماید.

### مواد و روش تحقیق

در این پژوهش، عناصر تأثیرگذار روی غلظت آلاینده‌ها در ۴ دسته عوامل ترافیکی، غلظت آلاینده‌ها در روزهای قبلی، داده‌های هواشناسی و عوامل مکانی طبقه‌بندی شده و مورد استفاده قرار می‌گیرند. داده‌های هواشناسی مورد استفاده شامل دما، رطوبت نسبی، سرعت باد، جهت باد، پوشش ابر و

<sup>۲</sup>Random forest feature selection

<sup>۱</sup>Wavelet transformation

در این پژوهش از تجزیه موجک، به‌عنوان یک روش پیش‌پردازش به‌منظور تجزیه داده‌های ورودی غلظت آلاینده‌ها در روزهای قبل استفاده شد. در این تحقیق، پس از انجام آزمایش‌های مقدماتی بر روی مجموعه داده‌ها و استفاده از توابع مختلف موجک، تابع موجک DB<sup>۲</sup>، به دلیل ارائه کوچک‌ترین تغییرات سیگنال در سطوح مختلف انتخاب گردید. چالش دیگر در تبدیل موجک تعیین بیشینه سطح ز، در اجرای این تبدیل می‌باشد. در سطوح بالاتر، تنوع سیگنال کمتر می‌باشد و بنابراین پیش‌بینی آن آسان‌تر است و انتظار می‌رود صحت بهتر باشد. در این تحقیق، مقدار J با استفاده از انحراف معیار سیگنال کلیات در همان سطح تعیین شده است. تجزیه سیگنال در سطحی که انحراف معیار سیگنال کلیات به‌صورت قابل ملاحظه‌ای کوچک‌تر از سیگنال اصلی باشد، متوقف خواهد شد. این شرایط با استفاده از رابطه تجربی ۳ بیان شده است (Siwek & Osowaski, 2012:1246):

$$\frac{std(A_j)}{std(x)} < 0.1 \quad \text{رابطه (۳)}$$

با اجرای رابطه ۳، مقدار  $\frac{std(A_j)}{std(x)}$  ۰/۰۵ به دست آمد و باتوجه به آن مقدار J=۵ محاسبه گردید. بدین ترتیب به‌منظور ورود داده‌ها به شبکه عصبی و روش رگرسیون چندمتغیره، مقادیر بیشینه و میانگین غلظت هر آلاینده برای هرروز در طی سال در ۵ سطح محاسبه گردید. بنابراین ورودی‌های شبکه عصبی با استفاده از تبدیل موجک مقادیر سیگنال جزئیات در ۵ سطح و سیگنال کلیات در سطح پنجم برای مقادیر غلظت آلاینده موردنظر برای هرروز و مقادیر میانگین سایر متغیرها می‌باشد. به‌منظور ارزیابی عملکرد مدل‌های پیش‌بینی، شاخص‌های ریشه میانگین مربعات خطا (RMSE)؛<sup>۴</sup> خطای استاندارد تخمین (SSE)؛<sup>۵</sup> ضریب تعیین (R<sup>2</sup>) و درصد خطا محاسبه شدند.

آزمون تی نمونه‌های جفت شده<sup>۶</sup> جهت ارزیابی آماری مدل‌های به‌دست‌آمده و مقایسه عملکرد مدل‌ها با یکدیگر مورد استفاده قرار گرفت. همچنین آزمون تی دو نمونه مستقل، با استفاده از مقادیر بیشتر و کمتر از میانگین داده‌های آزمایشی بهترین مدل‌ها در هر آلاینده به‌منظور تحلیل خطای این مقادیر انجام پذیرفت.

### محدوده مورد مطالعه

در سال‌های اخیر، آلودگی هوای تهران به‌عنوان یک چالش جدی مطرح شده است تا جایی که آلودگی هوای تهران در روزهای زیادی از سال، بیشتر از استانداردهای کیفیت هوا

قسمت داده‌ها و همچنین میانگین و بیشینه غلظت آلاینده موردنظر تا ۴ روز قبل، به‌عنوان ورودی مدل‌ها و ماکزیمم غلظت آلاینده‌ها در هر روز به‌عنوان خروجی مدل معرفی گردیدند. در این پژوهش مجموعه داده‌ها شامل ۱۷۸۰ نمونه از میانگین غلظت ساعتی برداشت شده از ۵ ایستگاه پایش در شهر تهران می‌باشد. مجموعه داده‌ها به‌منظور آموزش و آزمایش مدل‌ها و ایجاد یک مجموعه برای ارزیابی به‌منظور تنظیم پارامترهای مدل‌ها به صورت تصادفی به سه زیرمجموعه شامل ۷۰ درصد برای آموزش، ۱۵ درصد برای اعتبارسنجی پارامترهای مدل شبکه عصبی و یافتن پارامترهای بهینه، و ۱۵ درصد برای آزمایش مدل تقسیم شدند. جهت استفاده از روش رگرسیون چندمتغیره گام‌به‌گام، مسئله هم‌خطی چندگانه مدل‌ها مورد توجه قرار گرفت. به همین منظور شاخص شرطی<sup>۷</sup> و رواداری<sup>۸</sup> و نیز پی - مقدار<sup>۹</sup> برای هر یک از مدل‌ها، محاسبه گردید و مدل‌هایی که دارای شاخص شرطی بزرگ‌تر از ۳۰، مقدار رواداری کمتر از ۰/۱ و یا پی - مقدار بیشتر از ۰/۵ بودند به‌عنوان مدل‌های تحت تأثیر هم‌خطی چندگانه شناسایی شده و حذف گردیدند (Shamsoddini et al., 2015:1995). همچنین جهت یافتن تعداد نورون‌های مناسب، و دیگر پارامترها شامل تعداد تکرار، آستانه خطا، نرخ آموزش، و مومنتوم شبکه چندین بار اجرا شد و مقادیر بهینه برای هر آلاینده استخراج گردید. روش انتخاب خصیصه جنگل تصادفی به‌منظور انتخاب خصیصه‌های مناسب پیش از اجرای روش‌های مدل‌سازی مورد استفاده قرار گرفت (Shamsoddini et al., 2014). برای یافتن تعداد مناسب درخت برای هر جنگل تصادفی، تعداد تکرارها ۱۰، ۵۰، ۱۰۰ و ۲۰۰ تعیین شد. پس از انتخاب خصیصه‌های مناسب برای هر تعداد تکرار، خصیصه‌های انتخاب شده برای محاسبه صحت پیش‌بینی هر مجموعه از خصیصه‌ها، اعتبارسنجی گردید. مجموعه خصیصه‌هایی که بیشترین صحت پیش‌بینی را در مجموعه اعتبارسنجی دارا بودند به‌عنوان خصیصه‌های مناسب انتخاب گردیدند (Shamsoddini et al., 2014). پس از انتخاب ویژگی‌های مهم برای هر آلاینده، مدل‌های پیش‌بینی آلاینده‌ها با استفاده از ویژگی‌های انتخاب شده و روش‌های شبکه عصبی و رگرسیون چندمتغیره بدون استفاده از روش انتخاب ویژگی گام‌به‌گام، ایجاد گردیدند.

<sup>۴</sup>Standard Error Of Estimation

<sup>۵</sup>Correlation Of Determination

<sup>۶</sup>Paired-Samples T-test

<sup>۷</sup>Condition Index

<sup>۸</sup>Tolerance

<sup>۹</sup>P-Level

<sup>۱۰</sup>Root-Mean-Square Error



می‌باشد (Akbari et al., 2015:3551). در شکل ۱ منطقه مورد مطالعه و ایستگاه‌های سنجش آلاینده‌های هوا نشان داده شده است.



شکل ۱- موقعیت جغرافیایی منطقه مورد مطالعه و مکان‌های پایش آلودگی - (منبع: یافته‌های پژوهش، ۱۴۰۰).

### بحث و یافته‌های تحقیق

ماه در سال نیز در پیش‌بینی این آلاینده نقش مؤثری را ایفا می‌نماید. همان گونه که نتایج جدول ۱ نشان می‌دهد، متغیرهای بیشینه و میانگین غلظت در روزهای قبل به منظور پیش‌بینی آلاینده دی‌اکسید نیتروژن دارای بیشترین اهمیت می‌باشند. همچنین، متغیرهای بیشینه و میانگین غلظت در روزهای قبل دارای بیشترین تأثیر به منظور پیش‌بینی آلاینده دی‌اکسید گوگرد می‌باشند. همان طور که در جدول ۱ دیده می‌شود همانند دیگر آلاینده‌ها، آلاینده  $PM_{2.5}$  نیز به مقادیر غلظت در روزهای قبل وابستگی زیادی را نشان می‌دهد. همچنین وجود متغیر میانگین دما، نشان‌دهنده تأثیر قابل توجه آن در پیش‌بینی آلاینده  $PM_{2.5}$  می‌باشد.

به منظور استخراج تأثیرگذاری هر یک از متغیرها در پیش‌بینی آلاینده‌ها و تشخیص پارامترهای مؤثر در مدل‌سازی آلاینده‌ها، از روش جنگل تصادفی استفاده گردید که نتایج آن در جدول ۱ آورده شده است. همان طور که مشاهده می‌شود برای همه آلاینده‌ها پارامتر غلظت در روزهای قبل بیشترین میزان اهمیت در پیش‌بینی غلظت آلاینده‌ها را دارا می‌باشد. باتوجه به جدول ۱ مؤثرترین متغیر در پیش‌بینی غلظت منو اکسید کربن، بیشینه غلظت در یک روز قبل و بعد از آن میانگین غلظت در یک روز قبل و بیشینه غلظت در دو روز قبل می‌باشند. همچنین شاخص

جدول ۱- نتایج به دست آمده از روش جنگل تصادفی برای میزان اهمیت هر خصیصه در پیش‌بینی هر آلاینده به درصد

VIF( $PM_{2.5}$ )	VIF( $NO_2$ )	VIF( $SO_2$ )	VIF( $CO$ )	پارامتر
۲۶/۷	۲۶	۲۶/۲	۲۰/۳	بیشینه غلظت یک روز قبل
۲۳/۱	۲۰/۹	۱۷/۶	۱۳/۷	میانگین غلظت یک روز قبل
۴/۶	۷/۶	۶/۴	۴/۸	بیشینه غلظت دو روز قبل
۲/۴	-	-	۴/۷	شاخص ماه در سال
۷/۴	۶/۹	-	-	میانگین غلظت دو روز قبل
۲/۱	۶/۲	-	-	میانگین غلظت چهار روز قبل
۳	۵/۳	-	-	میانگین غلظت سه روز قبل
۲/۷	۴	-	-	بیشینه غلظت سه روز قبل
-	۴	-	-	بیشینه غلظت چهار روز قبل
۴/۸	-	-	-	میانگین دما

(منبع: یافته‌های پژوهش، ۱۴۰۰).

است. همچنین با استفاده از این مدل می‌توان گفت که ۹۰ درصد از تغییرات آلاینده  $PM_{2.5}$  با استفاده از متغیرهای مورد استفاده قابل پیش‌بینی می‌باشد که حاکی از کارایی این مدل برای پیش‌بینی آلاینده مورد بررسی است. نتایج حاصل از آزمون تی برای مقایسه عملکرد روش‌های مورد استفاده در پیش‌بینی آلاینده‌های مورد استفاده در این پژوهش در جدول ۳ نشان داده شده است. نتایج جدول ۳ نشان می‌دهد که در تمامی آلاینده‌ها تفاوت عملکرد روش MLP/Wavelet نسبت به سایر روش‌ها، در سطح آماری ۵ درصد معنی‌دار می‌باشد.

همچنین همان گونه که عنوان گردید، داده‌های مربوط به هر آلاینده بر اساس میانگین اندازه‌گیری‌های این داده‌ها به دودسته شامل نمونه‌هایی که مقدار آلاینده در آن‌ها کمتر از میانگین است و نمونه‌هایی که مقدار آلاینده آن‌ها مساوی یا بزرگ‌تر از میانگین مقادیر آن آلاینده می‌باشد، تقسیم شدند. سپس آزمون تی نمونه‌های مستقل بر روی مقادیر باقیمانده حاصل از تفریق داده‌های اندازه‌گیری شده و داده‌های پیش‌بینی شده بکار گرفته شد. نتایج حاصل از آزمون تی نمونه‌های مستقل در سطح معنی‌داری ۵ درصد برای بهترین روش‌های پیش‌بینی هر آلاینده نشان داد که، برای آلاینده‌های منو اکسید کربن، دی اکسید نیتروژن و دی اکسید گوگرد مقادیر بالاتر از میانگین با خطای بیشتری تخمین زده می‌شود. به عبارت دیگر در زمانی که مقدار این آلاینده‌ها در هوا زیاد باشد، صحت پیش‌بینی این آلاینده‌ها کاهش می‌یابد. همچنین برای آلاینده  $PM_{2.5}$  از لحاظ آماری تفاوت معناداری بین مقادیر بالاتر و پایین‌تر از میانگین مشاهده نگردید. به عبارت دیگر در پیش‌بینی این آلاینده، مقدار آلاینده در هوا نمی‌تواند در میزان خطای پیش‌بینی تأثیرگذار باشد.

جدول ۲- نتایج حاصل از به‌کارگیری روش‌های رگرسیون چندمتغیره و شبکه عصبی پرسپترون چندلایه بر روی داده‌های آزمایشی جهت پیش‌بینی آلاینده‌ها

آلاینده	روش	ضریب تعیین	ریشه میانگین مربعات خطا	خطای استاندارد تخمین	خطای استاندارد
آلاینده منو اکسید کربن (PPM) <sup>۲</sup>	SMLR	۰/۵۶	۳/۲	۳۳/۲	۴۳/۵۰
	MLR/RandomForest	۰/۴۸	۳۸/۱	۳۹/۱	۳/۳۰
	SMLR/Wavelet	۰/۸۵	۰/۸	۰/۹۵	۲۱/۱۶
	MLP	۰/۵۳	۴۱/۱	۵۲/۱	۹/۳۲
	MLP/RandomForest	۰/۴۷	۳۹/۱	۳۹/۱	۳۹/۳۰
آلاینده دی اکسید نیتروژن (PPB) <sup>۳</sup>	MLP/Wavelet	۰/۸۶	۰/۷۵	۰/۸۹	۱۹/۸۲
	SMLR	۰/۷۴	۵۴/۱۵	۶۶/۱۵	۱۶/۲۷
	MLR/RandomForest	۰/۷۴	۲/۱۲	۲۷/۱۲	۸۲/۲۲
	SMLR/Wavelet	۰/۱۶	۲۳/۱۵	۲۷/۵۲	۵۱/۷۹
	MLP	۰/۷۴	۱۴/۹۲	۱۲/۱۵	۹/۲۶
آلاینده دی اکسید نیتروژن (PPB) <sup>۳</sup>	MLP/RandomForest	۰/۷۶	۱۲/۰۴	۲۲/۱۲	۷۲/۲۲
	MLP/Wavelet	۰/۷۱	۷/۶۵	۹/۰۹	۱۷/۱۱

<sup>۲</sup> - Parts per billion

<sup>۳</sup>Independent Samples T-Test  
<sup>۳</sup>Parts per million

ادامه جدول ۲- نتایج حاصل از به‌کارگیری روش‌های رگرسیون چندمتغیره و شبکه عصبی پرسپترون چندلایه بر روی داده‌های آزمایشی جهت پیش‌بینی آلاینده‌ها

خطای استاندارد	خطای استاندارد تخمین	ریشه میانگین مربعات خطا	ضریب تعیین	روش	آلاینده
۵۷/۳۹	۹۶/۱۳	۱۲/۹۳	۰/۶۱	SMLR	آلاینده دی‌اکسید گوگرد (PPB)
۶۵/۳۵	۱۲/۱۳	۱۳/۰۴	۰/۵۹	MLR/RandomForest	
۲۵/۹۳	۹/۲۴	۷/۷۷	۰/۸۶	SMLR/Wavelet	
۳/۳۹	۷۷/۱۳	۱۲/۷۵	۰/۶۱	MLP	
۹۵/۳۴	۸۸/۱۲	۱۲/۸۱	۰/۶۱	MLP/RandomForest	
۱۹/۵۶	۶/۹۷	۵/۸۷	۰/۹۲	MLP/Wavelet	آلاینده PM <sub>2.5</sub> (PPB)
۲۵/۳۵	۷۷/۲۰	۱۸/۷۷	۰/۴۲	SMLR	
۵۱/۳۲	۱۵/۲۰	۱۹/۸۵	۰/۴۲	MLR/RandomForest	
۱۹/۹	۱۲/۲۳	۱۰/۲۹	۰/۸۶	SMLR/Wavelet	
۴۶/۳۳	۷۲/۱۹	۱۸/۳۵	۰/۴۶	MLP	
۳۲/۳۲	۲۴/۱۹	۱۸/۱۳	۰/۴۹	MLP/RandomForest	
۱۷/۰۴	۱۰/۴۷	۸/۸۱	۰/۹	MLP/Wavelet	

(منبع: یافته‌های پژوهش، ۱۴۰۰).

توانسته است تا مدل بهتری را جهت پیش‌بینی این آلاینده ارائه نماید. با استفاده از تبدیل موجک، بجای پیش‌بینی غلظت کل آلاینده، مدل‌ها به پیش‌بینی ضرایب موجک در سطوح مختلف می‌پردازند. به عبارت دیگر بازسازی غلظت نهایی آلاینده بر اساس ضرایب موجک با استفاده از عملیات ساده جمع انجام می‌پذیرد (Osowski & IouGaranty, 2007:745). خطای برآورد ۵۰/۴۳ در صد که با استفاده از مدل رگرسیون چندمتغیره به دست آمده است حاکی از روابط غیرخطی بین فضای ورودی و خروجی برای آلاینده منو اکسید کربن می‌باشد. با توجه به نتایج به دست آمده، خطای حاصل از مدل‌های MLP و MLR/Random Forest و MLP/Random Forest تفاوت بسیار اندکی را نشان می‌دهد که نشانگر کارایی اندک روش انتخاب ویژگی با استفاده از جنگل تصادفی و کاهش ابعاد داده‌های ورودی در رابطه با مدل‌سازی این آلاینده می‌باشد. همچنین تفاوت اندکی بین مدل‌های خطی و غیرخطی در ترکیب با روش انتخاب ویژگی با استفاده از جنگل تصادفی وجود دارد.

بر اساس نتایج این تحقیق، غلظت آلاینده‌ها در روزهای قبل به‌عنوان مهم‌ترین ویژگی‌ها به‌منظور پیش‌بینی مقادیر آلاینده‌ها در روز بعد انتخاب شدند. انتخاب ویژگی‌های مربوط به غلظت آلاینده‌ها در روزهای قبل می‌تواند نشان‌دهنده پایداری بالای این آلاینده‌ها در جو باشد که متعاقباً موجب می‌گردد تا غلظت آلاینده‌ها در روزهای قبل بر مقدار آلاینده در روزهای بعد تأثیر گذار باشد (Osowski & IouGaranty, 2007:745). شاخص ماه در سال برای آلاینده‌های CO و PM<sub>2.5</sub> به‌عنوان یکی از ویژگی‌های مهم انتخاب گردید. انتخاب شاخص ماه در سال نشان‌دهنده اهمیت تغییرات در تابش خورشیدی و شرایط جو در طول سال و تأثیر آنها بر میزان غلظت آلاینده‌ها می‌باشد (Arhami et al., 2013:4777). همچنین وجود متغیر دما در پیش‌بینی آلاینده PM<sub>2.5</sub> می‌تواند باعث افزایش انرژی جنبشی ذرات و حرکت عمودی این آلاینده شده و بر گسترش آلاینده در محیط تأثیر گذار باشد. به نظر می‌رسد با توجه به تغییرات زیاد غلظت آلاینده منو اکسید کربن با واریانس ۴۰۰۳۸۸۴، استفاده از تبدیل موجک با توجه به استخراج بخش‌های مفید داده‌ها و کاهش پیچیدگی محاسباتی،

جدول ۳- نتایج حاصل از آزمون تی نمونه‌های جفت شده جهت مقایسه روش‌های مورد استفاده

روش	CO	NO2	SO2	PM2.5
MLR و MLR/Random Forest	۰/۰۰	۰/۰۰	۰/۶۳	۰/۵۶
MLR و MLP	۰/۰۰	۰/۷۲	۰/۳۷	۰/۵۱
MLP و MLP/Random Forest	۰/۸۶	۰/۰۰	۰/۵۹	۰/۶۷
MLR و MLP/Random Forest	۰/۰۰	۰/۰۰	۰/۸۷	۰/۹
MLP و MLR/Random Forest	۰/۷۷	۰/۰۰	۰/۸۹	۰/۳۹
MLR Random Forest و MLP/Random Forest	۰/۰۹	۰/۸۸	۰/۰۳	۰/۳۳
MLR و MLR/Wavelet	۰/۰۰	۰/۰۱	۰/۰۰	۰/۰۳
MLP و MLP/Wavelet	۰/۰۰	۰/۰۰	۰/۰۰	۰/۰۰
MLR و MLP/Wavelet	۰/۰۰	۰/۰۰	۰/۰۰	۰/۰۰
MLP و MLR/Wavelet	۰/۰۲	۰/۰۰	۰/۰۰	۰/۰۰
MLR Wavelet و MLP/Wavelet	۰/۱	۰/۰۰	۰/۰۳	۰/۰۴

(منبع: یافته‌های پژوهش، ۱۴۰۰).

روش‌های خطی و غیرخطی در ترکیب با روش جنگل تصادفی برای آلاینده PM<sub>2.5</sub> وجود دارد. باتوجه به نتایج ارائه‌شده در جدول ۳، مقادیر کمتر از ۰/۰۵ نشان‌دهنده اختلاف معنی‌دار بین روش‌ها در سطح معنی‌داری ۵ درصد برای پیش‌بینی هر آلاینده می‌باشد. برای آلاینده CO و NO<sub>2</sub>، ویژگی‌های انتخاب‌شده توسط روش انتخاب ویژگی با استفاده از جنگل تصادفی باعث بهبود صحت مدل‌سازی حاصل از رگرسیون چندمتغیره گردیده است. درحالی‌که استفاده از ویژگی‌های انتخاب‌شده در این روش به‌منظور مدل‌سازی با استفاده از شبکه عصبی بهبود قابل‌ملاحظه‌ای را در صحت پیش‌بینی‌ها نشان نمی‌دهد. عدم کارایی روش انتخاب ویژگی جنگل تصادفی می‌تواند به دلایل تعداد درخت‌ها، تعداد متغیرهای انتخاب‌شده در هر گره و حداقل اندازه گره‌های پایانی باشد (Ghafouri kesbi et al, 2016). برای روش غیرخطی، روش انتخاب ویژگی جنگل تصادفی فقط در آلاینده NO<sub>2</sub> باعث بهبود صحت پیش‌بینی این آلاینده گردیده است. باتوجه به نتایج حاصل از آزمون تی برای آلاینده CO، مقدار پی-مقدار ۰/۱ برای مدل‌های MLP/Wavelet و MLR/Wavelet، تفاوت بسیار اندکی بین این دو مدل مشاهده می‌گردد که نشان‌دهنده عدم تفاوت معنی‌دار در عملکرد روش‌های غیرخطی و خطی بعد از اعمال تبدیل موجک در مدل‌سازی این آلاینده می‌باشد. همچنین فقط در آلاینده CO تفاوت معنی‌داری بین روش‌های خطی و غیرخطی مشاهده می‌گردد. باتوجه به نتایج حاصل از روش آزمون تی نمونه‌های مستقل، برای آلاینده‌های منو اکسید کربن، دی‌اکسید نیتروژن و دی‌اکسید گوگرد، نتایج پیش‌بینی آلاینده‌ها برای نمونه‌هایی با مقادیر بالاتر از میانگین خطای بالاتری را نشان می‌دهد. مقادیر وار یانس برای آلاینده منو اکسید کربن (وار یانس پایین‌تر از

عملکرد بهتر مدل به دست آمده برای آلاینده دی‌اکسید نیتروژن با استفاده از شبکه عصبی پرسپترون چندلایه به همراه روش تبدیل موجک، نشان‌دهنده کارایی این مدل در کاهش بار محاسباتی شبکه و ساده‌سازی شبکه با استفاده از محاسبه ضریب‌های موجک با تغییرپذیری پایین‌تر نسبت به سیگنال اصلی با تغییرپذیری بالاتر باشد. نتیجه ضعیف‌تر روش MLR/Wavelet نسبت به دو روش خطی دیگر می‌تواند بیانگر ناتوانی این مدل در برقراری ارتباط بین ورودی‌ها و خروجی‌ها می‌باشد. همچنین تفاوت ناچیز بین مدل شبکه عصبی پرسپترون چندلایه و رگرسیون خطی چندمتغیره نشان‌دهنده انتخاب ورودی‌های مناسب برای این مدل‌ها و رابطه نه‌چندان پیچیده بین ورودی‌ها و غلظت این آلاینده می‌باشد. مطابق نتایج ارائه‌شده در جدول ۲ برای آلاینده دی‌اکسید گوگرد بهترین مدل برای پیش‌بینی این آلاینده با استفاده از شبکه عصبی پرسپترون چندلایه در ترکیب با روش تبدیل موجک، به دست آمد که می‌تواند ناشی از استخراج اطلاعات مفید از داده‌ها در سطح‌های مختلف با استفاده از ضرایب این روش باشد. عملکرد مدل رگرسیون چندمتغیره تقریباً مشابه مدل شبکه عصبی پرسپترون چندلایه می‌باشد و استفاده از روش شبکه عصبی پرسپترون چندلایه تأثیر قابل‌ملاحظه‌ای در پیش‌بینی آلاینده دی‌اکسید گوگرد نسبت به استفاده از روش رگرسیون خطی نداشته است. در مجموع روش انتخاب ویژگی با استفاده از جنگل تصادفی در ترکیب با روش‌های خطی و غیرخطی مورد استفاده در این تحقیق، باعث افزایش صحت پیش‌بینی آلاینده دی‌اکسید گوگرد نسبت به زمانی شده است که مدل‌ها به صورت بدون این روش پیش‌پردازش مورد استفاده قرار می‌گیرند. باتوجه به نتایج به دست آمده تفاوت ناچیزی بین



اکسید کربن نسبت به سایر آلاینده‌ها، می‌تواند تنوع و تغییرپذیری بالایی منابع ایجادکننده این آلاینده مانند حمل‌ونقل منطقه‌ای و فرامنطقه‌ای و منابع ترافیکی و غلظت بالایی انباشته شده آلاینده در مناطق شهری مرتبط باشد. عامل دیگر باتوجه به واریانس بالایی غلظت این آلاینده در طول سال، می‌تواند ناشی از رفتار تصادفی (وجود نویز بیشتر) و نوسان بیشتر غلظت این آلاینده نسبت به سایر آلاینده‌ها باشد. علاوه بر این دلیل، عدم توانایی پارامترهای ورودی مدل در بازسازی فرایندهای اکسیداسیون تشکیل شده در جو (Arhami et al., ۲۰۱۳:۴۲۲) می‌تواند دلیل دیگری بر عدم دستیابی به صحت بالا برای پیش‌بینی غلظت منو اکسید کربن در این تحقیق باشد. در پایان باید به این نکته اشاره نمود که علاوه بر ماهیت آلاینده‌ها و نیز عملکرد مدل‌های پیش‌بینی‌کننده و نیز روش‌های پردازش خصیصه‌های پیش‌بینی‌کننده که در این تحقیق بدان‌ها پرداخته شد، عوامل تأثیرگذار دیگری نیز در پیش‌بینی آلاینده‌های هوا می‌توانند نقش داشته باشند. صحت پیش‌بینی غلظت آلاینده، بستگی به صحت اندازه‌گیری داده‌ها، منابع انتشار و تنوع فضایی غلظت آلاینده دارد (Dunea et al., ۲۰۱۵:۱).

باتوجه به پارامترهای استفاده‌شده برای مدل‌سازی غلظت آلاینده‌ها با روش جنگل تصادفی، پارامترهای غلظت در روزهای قبل از اهمیت زیادی برای پیش‌بینی آلاینده‌ها برخوردارند.

در زمان عدم استفاده از روش‌های تبدیل موجک و انتخاب ویژگی جنگل تصادفی، در پیش‌بینی آلاینده  $PM_{2.5}$ ، دی‌اکسید گوگرد و دی‌اکسید نیتروژن، تفاوت معنی‌داری بین مدل‌های خطی و غیرخطی وجود ندارد.

نتایج پیش‌بینی غلظت آلاینده منو اکسید کربن نسبت به آلاینده‌های دیگر از صحت پایین‌تر و آلاینده  $PM_{2.5}$  از صحت بالاتری برخوردار بود.

در زمانی که غلظت آلاینده‌های  $CO$ ،  $NO_2$  و  $SO_2$  در هوا بالاست، پیش‌بینی غلظت این آلاینده‌ها با خطای بیشتری توأم می‌باشد، درحالی‌که برای پیش‌بینی مقادیر غلظت پایین این آلاینده‌ها، دستیابی به صحت بالاتری انتظار می‌رود.

## References

۱- Ahadi, M., Sajadi, Zh., Yarigholi, V., (2019): Analysis and evaluation of livability indicators in urban areas Case study: 34 districts of Zanjan. Journal of Regional planning, 9(34), 131-148(In Persian).

میانگین برابر با  $0.51$  و بالاتر از میانگین برابر با  $3/38$ ، آلاینده دی‌اکسید نیتروژن (واریانس پایین‌تر از میانگین برابر با  $84$  و بالاتر از میانگین برابر با  $265$ ) و آلاینده دی‌اکسید گوگرد (واریانس پایین‌تر از میانگین برابر با  $57$  و بالاتر از میانگین برابر با  $277$ ) می‌تواند دلیل صحت کمتر پیش‌بینی‌ها برای مقادیر بالاتر از میانگین باشد. به عبارت دیگر مقدار بالایی واریانس برای نمونه‌های با مقادیر بالاتر از میانگین و متعاقباً پیچیدگی بیشتر این داده‌ها، می‌تواند علت اصلی صحت بالاتر پیش‌بینی غلظت آلاینده‌ها با مقادیر پایین‌تر از میانگین باشد.

باتوجه به نتایج آزمون‌های آماری برای آلاینده‌های مورد بررسی در این تحقیق، پیش‌بینی آلاینده منو اکسید کربن نسبت به سایر آلاینده‌ها با خطای بیشتری توأم می‌باشد. همچنین آلاینده  $PM_{2.5}$ ، نسبت به سایر آلاینده‌ها با صحت بالاتری پیش‌بینی می‌شود. روابط ساده‌تر و خطی بین پارامترهای ورودی مدل و آلاینده  $PM_{2.5}$  می‌تواند به صحت بالاتر پیش‌بینی این آلاینده نسبت به دیگر آلاینده‌ها منجر شود. همچنین تغییرات آلاینده منو اکسید کربن و عوامل تأثیرگذار بر این تغییرات نسبت به تغییرات دیگر آلاینده‌ها و عوامل تأثیرگذار بر تغییرات آن‌ها پیچیدگی بیشتری دارند. صحت پایین پیش‌بینی غلظت منو

## نتیجه‌گیری و ارائه پیشنهادها

در این پژوهش با استفاده از مدل‌های خطی و غیرخطی و روش‌های انتخاب ویژگی جنگل تصادفی و تبدیل موجک، به مدل‌سازی حداکثر غلظت آلاینده‌های منو اکسید کربن، دی‌اکسید نیتروژن، دی‌اکسید گوگرد و  $PM_{2.5}$  با استفاده از خصیصه‌های هواشناسی، غلظت آلاینده، خصیصه‌های مکانی و ترافیکی پرداخته شد. داده‌های هواشناسی شامل دما، رطوبت نسبی، سرعت و جهت باد، پوشش ابر و فشار سطح به صورت میانگین روزانه و داده‌های مکانی شامل مختصات ایستگاه‌ها و میانگین پوشش گیاهی تا شعاع یک کیلومتری بودند. نتایج حاصل از این تحقیق را می‌توان به صورت زیر عنوان نمود:

• مدل‌سازی غلظت آلاینده‌ها با استفاده از شبکه عصبی پرسپترون چندلایه به همراه تبدیل موجک حتی به مراتب بالاتر از نسبت به مدل‌های دیگر ارائه نمود.

۲- Ahadnejad, P., Khaledi, Sh., Ahmadi, M., (2020): Investigating the Long-term effect of dust on Health in order to prevent Its Impacts in Future Planning Case Study: Khuzestan

- Province. Journal of Regional planning, 10(39), 33-36(In Persian).
- ۳- Akbari, M., & Samadzadegan, F., (2015): Identification of air pollution patterns using a modified fuzzy co-occurrence pattern mining method. Int. J. Environ. Sci. Technol, 12, 3551-3562.
- ۴- Antanasijević, V. Pocajt, D. Povrenović, M., Ristić, A. Perić-Grujić., (2013): PM10 emission forecasting using artificial neural networks and genetic algorithm input variable optimization, Science of the Total Environment, Vol. 443, pp. 511-519.
- ۵- Arhami, M., Kamali, N., Rajabi, M., (2013): Predicting hourly air pollutant levels using artificial neural networks coupled with uncertainty analysis by Monte Carlo simulations. Environ Sci Pollut Res, 20, 4777-4789.
- ۶- Barrero, M.A., Grimalt, J.O., Canton, L., (2006): Prediction of daily ozone concentration maxima in the urban atmosphere. Chemom. Intell. Lab. Sys. 80, 67-76.
- ۷- Dunea, D., Pohoata, A., Iordache, S., (2015): Using wavelet-feedforward neural networks to improve air pollution forecasting in urban environments. Environ Monit Assess, 187(7), 1-16.
- ۸- Durao, M., Mendes, T., Pereira, M., (2016): Forecasting O3 levels in industrial area surroundings up to 24 h in advance, combining classification trees and MLP models. Atmospheric Pollution Research, 7, 961-970.
- ۹- Esmailnejad, M., Eskandari Sani, M., Borzaman, S., (2015): Evaluation and Zoning of Urban air Pollution in Tabriz. Journal of Regional planning, 5(19), 173-186(In Persian)
- ۱۰- Feng, X., Li, Q., Zhu, Y., Hou, J., Jin, L., Wang, J., (2015): Artificial neural networks forecasting of PM2.5 pollution using air mass trajectory based geographic model and wavelet transformation. Atmospheric Environment, 107, 118-128.
- ۱۱- Fernando, H. J. S., Mammarella, M. C., Grandoni, C., Fedele, P., Di Marco, R., Dimitrova, R., Hyde, P., (2012): Forecasting PM10 in metropolitan areas: efficacy of neural networks. Environ. Pollut. 163, 62-67.
- ۱۲- Ghafouri Kesbi, F., Rahimi Mianji, G., Honarvar, M., Nejati Javaremi, A., (2016): Tuning and Application of Random Forest Algorithm in Genomic Evaluation. Research on Animal Production, 7(13), spring and Summer(In Persian).
- ۱۳- Grivas, G., & Chaloulakou, A., (2006): Artificial neural network models for prediction of PM10 hourly concentrations, in the Greater Area of Athens, Greece. Atmospheric Environment, 40, 1216 - 1229.
- ۱۴- Karatzas, K. D., & Kaltsatos, S., (2007): Air pollution modelling with the aid of computational intelligence methods in Thessaloniki, Greece. Simulation Modelling Practice and Theory, 15, 1310-1319.
- ۱۵- Lee, S., Ho, CH., Choi, YS, (2011): High-PM10 concentration episodes in Seoul, Korea: background sources and related meteorological conditions. Atmos Environ, 45(39), 7240-7247.
- ۱۶- Mohammadi, N., Khatibi, KH., Shaker khatibi, M., Fatehi far, E., (2016): Predicting the concentration of gaseous pollutants in the air of Tabriz using a neural network. Civil and Environmental Engineering, 83(46) (In Persian).
- ۱۷- Moustris, K. P., Larissi, I. K., Nastos, P. T., Koukouletsos, K. V., Paliatsos, A. G., (2013): Development and Application of Artificial Neural Network Modeling in Forecasting PM10 Levels in a Mediterranean City. Water Air Soil Pollut, 224(8), 1634-1642.
- ۱۸- Noorani, V., Karimzadeh, H., Najafi, H., Hosseini, A., (2019): Predicting the concentration of NO2 and SO2 pollutants in the air of Tabriz using artificial neural network and adaptive neural-fuzzy inference system and comparing the obtained results. International Conference on civil engineering ,architecture and urban planning(In Persian).
- ۱۹- Osowski, S., & Iougaranty, K., (2007): Forecasting of the daily meteorological pollution using wavelets and support vector machine. Engineering Applications of Artificial Intelligence, 20, 745 - 755.
- ۲۰- Perez, P., & Trier, A., (2011): Prediction of NO and NO2 concentrations near a street with heavy traffic in Santiago, Chile. Atmos. Environ., 35, 1783-1789.
- ۲۱- Sadr Mousavi, M.S., & Rahimi, A., (2010): Comparison of Multilayer Perceptron Neural Networks with multiple regression to predict the concentration of ozone in Tabriz, Natural Geography Research, Vol. 71, pp. 65-72(In Persian).
- ۲۲- Shamsoddini, A., Raval, S., Taplin, R., (2014): Spectroscopic analysis of soil metal contamination around a derelict mine site in the blue mountains, australia”, ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial

- Information Sciences, II-7, 2014 ISPRS Technical Commission VII Symposium, 29 September – 2 October 2014, Istanbul, Turkey.
- ۲۳- Shamsoddini, A., Trinder, J. C., Turner, R., (2015): Paired-data fusion for improved estimation of pine plantation structure. *International Journal of Remote Sensing*, 36, 1995-2009.
- ۲۴- Shamsoddini, A., Aboodi, M. R., Karami, J., (2017): Tehran air pollutants prediction based on random forest feature selection method. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-4/W4, 2017 Tehran's Joint ISPRS Conferences of GI Research, SMPR and EOEC 2017, 7–10 October 2017, Tehran, Iran.
- ۲۵- Sharma, M., Aggrawal, S., Bose, P., (2002): Meteorology – base forecasting of air quality index using neural network. *International conference neural network*, Singapore, 374-378.
- ۲۶- Siwek, K., & Osowaski, S. S., (2012): Improving the accuracy of prediction of PM10 pollution by the wavelet transformation and an ensemble of neural predictors. *Engineering Applications of Artificial Intelligence*, 25, 1246–1258.
- ۲۷- Tavakoli, M., & Esmaeili, A., (2014): Performance of ANN and fuzzy neural network adaptive for estimating of the concentration of suspended particles in the air of Tehran. *Journal of Environmental Science and Engineering*, 2, 75-84 (In Persian).
- ۲۸- Wang, P., Liu, Y., Qin, Z., Zhang, G., (2015): A novel hybrid forecasting model for PM10 and SO2 daily concentrations. *Science of the Total Environment*, 505, 1202–1212.
- ۲۹- Zhang, H., Zhang, W., Palazoglu, A., Sun, W., (2012): Prediction of ozone levels using a Hidden Markov Model HMM with Gamma distribution. *Atmos Environ*, 62, 64–73.



