Alexandria Engineering Journal 104 (2024) 464–479

Contents lists available at ScienceDirect

Alexandria Engineering Journal

journal homepage: www.elsevier.com/locate/aej

Machine learning models for predicting interactions between air pollutants in Tehran Megacity, Iran

Abdullah Kaviani Rad [a], Seraj-Odeen Razmi [b], Mohammad Javad Nematollahi [c,*], Armin Naghipour [d], Foroogh Golkar [e], Mohammadreza Mahmoudi [f,*]

[a] *Department of Environmental Engineering and Natural Resources, College of Agriculture, Shiraz University, Shiraz 71946-85111, Iran*
[b] *Department of MBA, Faculty of Management, University of Tehran, Tehran 14179-35840, Iran*
[c] *Department of Geology, Faculty of Sciences, Urmia University, Urmia 57561-51818, Iran*
[d] *Clinical Research Development Center of Imam Khomeini Hospital and Dr. Mohammad Kermanshahi and Farabi, Kermanshah University of Medical Sciences, Kermanshah 67148-69914, Iran*
[e] *Department of Water Engineering & Oceanic and Atmospheric Research Center, Shiraz University, Shiraz 71441-65186, Iran*
[f] *Department of Statistics, Faculty of Science, Fasa University, Fasa 74616-86131, Iran*

ARTICLE INFO

Keywords:
Air pollution
Air quality
Modeling
Machine learning
Simulation

ABSTRACT

Air pollution has significant detrimental impacts on the environmental compartments and human health. The present study investigates the interactions between selected air pollutants, including CO, O_3, NO_2, SO_2, PM_{10}, and $PM_{2.5}$, in Tehran megacity, the capital of Iran, using statistical modeling and simulating approaches. Data on selected air pollutants have been extracted from 16 sensors across Tehran during 2013–2022. Using the development of three machine learning models, including XGBoost (XGB), LightGBM (LGBM) and Random Forest (RF), significant relationships between the air pollutants were observed. The comparison of the models demonstrated that the RF model has the highest level of optimality in forecasting the concentrations of CO, O_3, NO_2, and SO_2 ($R^2_{CO} = 0.63$, $R^2_{O3} = 0.6$, $R^2_{NO2} = 0.54$, $R^2_{SO2} = 0.55$). The LGBM model provided higher optimality for NO_2 and $PM_{2.5}$ ($R^2_{NO2} = 0.56$, $R^2_{PM2.5} = 0.88$), while the XGB model exhibited higher accuracy for the PM_{10} ($R^2_{PM10} = 0.88$). Regarding the results obtained from the statistical models, it can be inferred that the RF model has superior performance in the forecasting of gaseous pollutants. Moreover, the XGB and LGBM models exhibited comparable performance and therefore may be regarded as appropriate options for PM prediction. The findings from the simulations indicated that a rise in an individual pollutant likely leads to an increase in other pollutants. Consequently, implementing air quality management strategies for a specific pollutant meaningfully, directly influences other pollutants, highlighting the significance of considering the chemical aspect of air pollutants interaction and enforcing air pollution management rules.

1. Introduction

Atmospheric pollution has emerged as a pressing global health threat [1–3]. Efforts to combat air pollution have proven relatively ineffective, and air quality is declining throughout many regions of the world [4]. Carbon monoxide (CO), nitrogen oxides (NO_X), ozone (O_3), sulfur dioxide (SO_2), and particulate matter (PM) are among the most frequently observed air contaminants [5]. Previous studies have demonstrated that air pollution has a negative impact on all phases of life as well as triggering both the onset and progression of numerous diseases [6,7]. Multiple epidemiological investigations have convincingly shown a significant association between air pollution exposure and higher rates of morbidity and mortality [5,8]. Air pollution contributes to approximately seven million premature mortalities annually, along with an elevated frequency of admissions to hospitals [9,10]. It is also expected to take responsibility for about 14 % of all stroke-related deaths, ranking it as the world's second-most prevalent driver of fatalities [11]. Air pollutants enhance the risk of ST-elevation myocardial infarction, sudden cardiac death, cardiac arrhythmia, and peripheral arterial disease [5,12,13]. Moreover, pollutants may play a role in mental disorders owing to their detrimental effects on the brain and nervous system. For instance, significantly elevated depression symptoms have been attributed to long-term exposure to $PM_{2.5}$ [14–16].

Moreover, a large number of studies have shown the impact of air

* Corresponding authors.
E-mail addresses: mj.nematollahi@urmia.ac.ir (M.J. Nematollahi), mahmoudi.m.r@fasau.ac.ir (M. Mahmoudi).

https://doi.org/10.1016/j.aej.2024.08.023
Received 7 May 2024; Received in revised form 29 June 2024; Accepted 6 August 2024
Available online 10 August 2024