Air Quality-Related Parameter Forecasting Using Soft Computing Techniques

Subjects: Engineering, Environmental

Contributor: Lakindu Mampitiya, Namal Rathnayake, Lee P. Leon, Vishwanadham Mandala, Hazi Md. Azamathulla, Sherly Shelton, Yukinobu Hoshino, Upaka Rathnayake

Air is one of the main factors that maintain the balance of the earth's environment. The World's Air Pollution: Real-Time Air Quality Index clearly indicates the recent air pollution levels of the earth. Machine learning algorithms such as random forest (RF), linear regression (LR), and support vector regression (SVR) play an essential role as traditional algorithms for regression purposes.

Keywords: air quality ; forecasting

1. Introduction

Air is one of the main factors that maintain the balance of the earth's environment. According to the World Health Organization (WHO), air pollution is one of the most significant environmental hazards. Anthropogenic activities are mainly responsible for the unbalance of the environment that has triggered air pollution. However, natural scenarios like climate change, volcanic eruptions, gases released from living creatures, and sea salt spray are some factors contributing to air pollution. Nevertheless, human activities have become the main contributing factor to air pollution over time. The most common human activities are burning fossil fuels and chemical manufacturing. Concerning those activities, nitrogen dioxide, sulfur dioxide, carbon dioxide, and carbon monoxide are the most common hazardous gases [1]. According to the WHO statistical analysis [2], around 2.4 billion people use open household fires with firewood, kerosene, and other biomass. This affects the reduction in air quality. This household emission was directly and indirectly responsible for 3.2 million deaths in 2020 [2]. Globally, the severity can be identified as 237,000 deaths of children under five years of age and 6.7 million premature deaths per year [2]. Ischemic heart disease, strokes, and lung cancers are also considered prominent effects in this case. Air quality is a statistical indicator that elaborates how the quality of the air varies, considering the available materialistic contents in the air. This factor shows all the possible ingredients in the air. Therefore, whole factors have to be considered for a holistic view of air quality. Different types of gases, various aerosols, and other particles are very common and impact air quality.

Currently, many countries are focusing on bringing their environmental air quality to a reasonable level by eliminating the emission of hazardous gases to the environment. Therefore, many countries have imposed laws and regulations to cut down the emissions level. Many European countries are moving to electric-powered vehicles by giving up fossil-fueled automobiles [3]. In addition, renewable energy generation has significantly increased over the last decade. These measures have helped reduce hazardous effects, not only in terms of air quality but also global warming, thus mitigating the impact of climate change. However, much has to be done to bring the atmosphere back normal, as it was before the Industrial Revolution. Therefore, many responsible agencies, including the WHO, have provided guidelines to minimize adverse impact on the atmosphere. Some of the WHO-published guidelines include keeping the atmosphere at the following levels:

- Ozone (O_3) → 100 μg/m^3 8-h mean;
- Nitrogen dioxide (NO_2) → 25 μg/m^3 24-h mean;
- Sulfur dioxide (SO_2) → 40 μg/m^3 24-h mean;
- Carbon monoxide (CO) → 7 μg/m^3 24-h mean;
- $PM_{2.5}$ → 15 μg/m^3 24-h mean.

The World's Air Pollution: Real-Time Air Quality Index clearly indicates the recent air pollution levels (as of March 2023) of the earth, which showcases that the air above Asian, African, and Latin American countries is not at a good level [3][4]. However, Asian countries are at a severe level as the majority of the world's population lives in those countries. Importantly, PM_{10} and $PM_{2.5}$ are at higher concentrations than what is given in WHO guidelines in most Asian countries. The situation is no different in Sri Lanka. The island has experienced some occasional severe air quality levels in 2022 and 2023 [5]. This is mainly due to emissions from vehicles, the burning of organic materials, power generation, and petroleum refining. However, on average, most of the areas have had moderate air quality conditions in 2023. The Air Quality Index (AQI) was within the range of 51 to 100.

2. Air Quality-Related Parameter Forecasting Using Soft Computing Techniques

Machine learning algorithms such as random forest (RF), linear regression (LR), and support vector regression (SVR) play an essential role as traditional algorithms for regression purposes. The functionality of the algorithms mainly depends on the dataset variations [6]. Saikiran et al. [6] conducted a study to predict air quality using machine learning algorithms using three algorithms, including RF, SVR, and LR. They achieved an RMSE value of 0.812 for RF as the highest functioning model for the specific dataset. In addition, Guo et al. [7] conducted a study to predict air quality using a limited amount of data (23 July 2020 to 13 July 2021) for Shanxi, China's meteorological station. The research study mainly focused on six parameters (SO_2, NO_2, PM_{10}, $PM_{2.5}$, O_3, and CO) in forecasting the AQI for a considered period range (8, 16, 24, 32, 40, and 48 h). As the functioning model, the researcher used an auto model network that could predict a functional capacity slightly above 50%. Importantly, the research team emphasized the model functionality with a comparative analysis [7], using state-of-the-art algorithms for the comparative analysis to verify the functionality of the proposed auto model.

However, Popa et al. [8] suggested that the optimized GPR algorithm outperforms other traditional algorithms, such as linear regression and SVM, in one of their research works carried out in Bucharest, Romania. Nevertheless, some other researchers stated that the hybrid deep learning models can function well in forecasting air quality data by competing with other traditional machine learning algorithms [9][10]. Therefore, it is well understood that algorithms can change their behavior based on the context of their usage, specifically with the location. Considering the comparative analysis, most of the research studies undertook this step to verify the proposed model [7][8][9][10][11][12][13]. The functionality of the models differs due to the dataset variation and the gaps between the data. Researchers have worked on gathering hourly data to identify every simple variation in changes in air quality [7][8][9][10][13]. The root mean squared error (RMSE), mean absolute error (MAE), mean squared error (MSE), mean absolute relative error (MARE), relative absolute error (RAE), and coefficient of determination (R^2) are some commonly used evaluating matrices to obtain a well-briefed evaluation on the models' performance [6][7][8][9][10][11][12][13].

On the other hand, forecasting AQI and related data is an important process that requires precise output. Because of that, consideration of more factors that have the capability to affect air quality may lead to a more reliable result. Many previous studies [6][7][8][9][10][11][12][13] have elaborated on the effect of considering such data on the final output of the models. It has been identified that most of the research studies considered CO, NO_2, O_3, SO_2, $PM_{2.5}$, and PM_{10} as their input parameters because of their importance.

References

1. Manisalidis, I.; Stavropoulou, E.; Stavropoulos, A.; Bezirtzoglou, E. Environmental and Health Impacts of Air Pollution: A Review. Front. Public Health 2020, 8, 14.
2. Air Pollution. Available online: https://www.who.int/health-topics/air-pollution#tab=tab_2 (accessed on 8 March 2023).
3. Soret, A.; Guevara, M.; Baldasano, J.M. The potential impacts of electric vehicles on air quality in the urban areas of Barcelona and Madrid (Spain). Atmos. Environ. 2014, 99, 51–63.
4. Imane, S.; Oumaima, B.; Kenza, K.; Laila, I.; Youssef, E.M.; Zineb, S.; Mohamed, E.J. A review on climate, air pollution, and health in North Africa. Curr. Environ. Health Rep. 2022, 9, 276–298.
5. Live Animated Air Quality Map (AQI, PM2.5...) | IQAir. Available online: https://www.iqair.com/air-quality-map?lat=7.61266509224&lng=80.7010823782&zoomLevel=7 (accessed on 18 February 2023).
6. Saikiran, K.; Lithesh, G.; Srinivas, B.; Ashok, S. Prediction of Air Quality Index Using Supervised Machine Learning Algorithms. In Proceedings of the ACCESS 2021-Proceedings of 2021 2nd International Conference on Advances in

Computing, Communication, Embedded and Secure Systems, Ernakulam, India, 2–4 September 2021; pp. 141–144.

7. Guo, Y.; Zhu, T.; Li, Z.; Ni, C. Auto-Modal: Air-Quality Index Forecasting with Modal Decomposition Attention. Sensors 2022, 22, 6953.
8. Popa, C.L.; Dobrescu, T.G.; Silvestru, C.I.; Firulescu, A.C.; Popescu, C.A.; Cotet, C.E. Pollution and Weather Reports: Using Machine Learning for Combating Pollution in Big Cities. Sensors 2021, 21, 7329.
9. Huang, C.J.; Kuo, P.H. A Deep CNN-LSTM Model for Particulate Matter (PM2.5) Forecasting in Smart Cities. Sensors 2018, 18, 2220.
10. Jang, J.; Shin, S.; Lee, H.; Moon, I.C. Forecasting the Concentration of Particulate Matter in the Seoul Metropolitan Area Using a Gaussian Process Model. Sensors 2020, 20, 3845.
11. Yu, R.; Yang, Y.; Yang, L.; Han, G.; Move, O.A. RAQ–A Random Forest Approach for Predicting Air Quality in Urban Sensing Systems. Sensors 2016, 16, 86.
12. Liaw, J.J.; Chen, K.Y. Using High-Frequency Information and RH to Estimate AQI Based on SVR. Sensors 2021, 21, 3630.
13. Wardana, I.N.K.; Gardner, J.W.; Fahmy, S.A. Optimising Deep Learning at the Edge for Accurate Hourly Air Quality Prediction. Sensors 2021, 21, 1064.