Engineering and Technology Journal for Research and Innovation (ETJRI)
ISSN 2581-8678, Volume III, Issue II, July 2021

Application of Geospatial Artificial Intelligence in mapping of air pollutants in urban cities

Sushree Sangita Mishra[1], Deepa Parasar[2*]
[1] *Associate Professor, CE, ASET, Amity University Maharashtra, ssmishra@mum.amity.edu*
[2*] *Associate Professor, CSE, ASET, Amity University Maharashtra, dparasar@mum.amity.edu*

Abstract— **People from most of the countries across the globe have been distressed due to air pollution induced acute health problems, which emphasize on the need of air pollution monitoring. Remote sensing of high resolution satellite imageries is a useful tool that facilitates a better interpretation and analysis of the concentration of air pollutants. Satellite imageries are helpful to quantify and map air pollution because they are able to provide synoptic view of large areas. Different satellite sensors with various ranges of temporal, spatial and radiometric resolution offer the opportunity to estimate the concentration of harmful air pollutants such as carbon monoxide, nitrogen dioxide, ammonia, sulphur dioxide, volatile organic compounds, and surface particulate matter (PM2.5). In addition to mapping of the air pollutants to check its detrimental effects on human health, it is also important to predict the future concentrations by developing models using different simulation methods. These simulated models may help the government officials and authorities for policy implementation towards the control of air pollution levels. Though many research works are available on different approaches to model air pollutant levels, recent advancements in deep learning techniques are more promising with accurate forecasting results. This paper presents a review of the remote sensing technique in mapping and prioritizing the concentration of air pollutants. In addition to this, a description of the deep learning architectures to develop models on air pollution data obtained from the satellite sensors is also discussed.**

Keywords: **Air pollution, Remote sensing, Deep learning, Sensors, Spatial interpolation.**

I. INTRODUCTION

Environmental pollution issues such as water, noise, and air pollution are emerging as cities grow economically and technologically. Air pollution, in particular, has a significant effect on human health due to the exposure to toxins and particulates, which have piqued the scientific community's interest in air pollution and its consequences. The combustion of fossil fuels, deforestation, waste from factories and industries, domestic heating, and natural disasters are the major sources of air pollution. Air pollution is one of the world's most critical issues. The consistency of the air varies dramatically over time and space. Because of the strong temporal resolution of different satellite sensors, satellite data can be used to quantify and map air pollution because it can provide full views of vast areas with just one image on a regular basis. Different air pollutants, such as nitrogen dioxide (NO2), sulphur dioxide (SO2), ammonia (NH3), carbon monoxide (CO), certain volatile organic compounds (VOCs), and aerosol optical depth (AOD), from which surface particulate matter is extracted, can be estimated using a range of satellites. To construct a surface grid or contour chart, spatial interpolation techniques are used. Using known concentrations at fewer points, interpolation processes estimate concentrations in the study region. A geographic information system (GIS) is a software tool for storing, analysing, managing, and presenting all forms of spatial or geographical data, which produces solutions to spatial problems when combined with the expert opinion of the GIS user or analyst. Various researchers have used GIS techniques to investigate the spatial and temporal distribution of contaminants. One of the aims of this project was to measure the temporal and spatial variation of air pollution across the country. An air quality monitoring network system is required to collect data on SO2, NO2, and SPM emissions concentrations. The point data files were also developed for the Mumbai region's air quality monitoring stations. At seasonal and annual time scales, air quality concentration data for SO2, NO2, and SPM was assigned to the appropriate point form file. At weekly, seasonal, and annual temporal scales, the collected air quality data set can be fed as an attribute of points form files. The current study's approach aims to discuss and review effective models that can be used to forecast current Particulate matter concentration levels. Such forecasts result in an estimation of air quality (in terms of Particulate Matter levels) that can be useful to local decision makers in charge of the day-to-day operational implementation of relevant legal structures, such as issuing pollution exceedance alerts and enforcing effective steps. Machine learning (ML) and Deep Learning models can be used to analyse the data obtained by RS and to meet the need for more accurate forecasting in various scientific areas and domains. ML algorithms, in particular, have been commonly used to predict air quality.

II. LITERATURE SURVEY

Owing to prolonged exposure to air contaminants such as nitrogen dioxide (NO2), particulate matter (PM), ozone (O3), and sulphur dioxide, millions of people across the world are suffering from severe health problems (SO2). According to the World Health Organization [10], air pollution killed over 4.2 million people in 2018.Furthermore, according to WHO, about 90% of the world's population lives in areas where air pollution levels exceed WHO air standards [10]. As a result, global attention was drawn to air pollution. Many countries have set up ground stations to track air quality, but the information gathered by these stations is inadequate to map air pollution because air quality is both spatially and temporally variable. Since satellite images have a high temporal resolution, they can cover a wide area in a single image, making satellite remote sensing a powerful tool for air pollution control [23].

Several studies have used satellite images to track and chart air contaminants. For example, OMI was used to monitor NOx emissions in Colombia [7], the mid-Atlantic United States [5], the United States [11] [8] [13] and Europe [16].OMI was also used to measure SO2 emissions in North America [9]. Abdel Satter [4] looked at the capabilities of various satellite remote sensing technologies for calculating air contaminants, as well as the methods for processing and accessing satellite data for pollutant concentration mapping. According to the researcher, different spectral resolutions of space instruments allow for the detection of various species of air pollutants. Air pollution measurements from space are useful for monitoring air quality and studying the long-term pattern of atmospheric pollutant concentrations. Support vector regression (SVR) [3] is used to predict pollutant and particulate levels and calculate the air quality index (AQI). Out of all the alternatives tested, the radial basis function (RBF) was the type of kernel that allowed SVR to make the most accurate predictions. The results showed that SVR with the RBF kernel could reliably predict hourly pollutant concentrations such as carbon monoxide, sulphur dioxide, nitrogen dioxide, ground-level ozone, and particulate matter 2.5, as well as the hourly AQI for California. Research was conducted [22] to create regression models that can predict the spatial and temporal variability of PM10 in cyprus (particulate matter with an aerodynamic diameter less than 10 m) ground measurements, Aerosol Optical

Depth retrievals from satellite, and the prevailing synoptic conditions developed by Artificial Neural Networks. An adequately developed neural classification system clustered synoptic charts, Aerosol Optical Depth data from the Aqua satellite, and ground measurements of particulate matter at the heart of the forecasting system. Statistical models for forecasting emission levels were created using the above tools. A similar study was conducted [24] in Equador to model PM2.5 urban emissions using machine learning and meteorological parameters. The classification model's findings indicate that low (10 g/m3) versus high (>25 g/m3) and low (10 g/m3) versus moderate (10–25 g/m3) PM2.5 concentrations are reliably classified. A regression analysis indicates a better estimate of PM2.5 when climatic conditions become more extreme (strong winds or high levels of precipitation). When climatic conditions become more severe, a regression analysis suggests a better estimate of PM2.5 (strong winds or high levels of precipitation).

Across the globe it has been identified that adoption of AI and machine learning in dealing with hazards of air pollution is on its frontiers. Deep learning has been described as a promising field in the situation, with the potential to increase the resolution and precision of predictions in data-poor scenarios, according to the literature. It has been argued that low- and middle-income countries with minimal monitoring and surveillance infrastructure may depend on image-based convolutional neural networks [21] as a cost-effective method of estimating local environmental exposures. Choosing Shanghai, China as the target city [2] combined two kinds of deep networks for extracting features in both spatial and temporal dimensions, Convolutional neural network is taken as the base layer, automatically extracting features of input data. A long short-term memory network is used for the output layer to consider the time dependence of pollutants in forecasting PM2.5 concentration. IoT [19] has been incorporated for measuring a variety of city parameters at Antwerp, Belgium, including air pollution. To deal with temporal and spatial resolution, the problem is addressed as matrix completion on graphs problem and relies on variation graph auto encoders to propose a deep learning solution for the estimation of the unknown air pollution values. A new deep learning method[17] was proposed based for air quality forecasting model, namely DFS (Deep Flexible Sequential) a hybrid of LSTM and CNN to forecast PM10 as the target pollutant through the work in the region of Istanbul Turkey between 2014 and 2018. A Deep Multi-output LSTM (DM-LSTM)[25] neural network model has been proposed by incorporating features of three deep learning algorithms (i.e., mini-batch gradient descent, dropout neuron and L2 regularization) to configure the model for extracting the key factors of complex spatio-temporal relations as well as reducing error accumulation and propagation in multi-step-ahead air quality forecasting at Taipei, City of Taiwan. This is advancement over the Shallow Multi-output Long Short-Term Memory (SM-LSTM) model which is suitable for regional multi-step-ahead air quality forecasting, but encounters spatio-temporal instabilities and time-lag effects.

III. METHODOLOGY

Geospatial artificial intelligence has been successfully used in many disaster related problems and agricultural studies by many researchers. It has also been used for air quality mapping and modelling of air pollutants in urban cities. This method encompasses two techniques viz. Geospatial Technology and artificial intelligence. The geospatial technology can be used for spatio- temporal mapping of air pollutants and prioritization of locations based on this analysis. The

Artificial intelligence techniques viz. Deep Learning techniques, ANN, CNN etc. are then applied to forecast the concentration of the pollutants to install air purifiers in the prioritized locations. A generalized methodology flowchart is mentioned below (Fig. 1).

A. Mapping of Air Pollutants

The concentration of air pollutants specifically SO2, NO2, and SPM are obtained from the either from the government organizations or can be extracted from satellite images using image processing software. The concentration of air pollutants are mapped and spatially interpolated using different GIS based spatial interpolation techniques viz. Kriging method, splines, Inverse distance weighted (IDW) method etc. These techniques have been used successfully in air pollution studies [12], [15]. IDW method is used when density of the set of points is high so that a local surface variation can be obtained for the analysis. It uses a linear weighted combination of data to compute the grid values of any random parameter. It depends on the distance between the sampling sites and the location where the interpolation is needed [14]. Points that are closer to the interpolation point for SPM would have more weight. The IDW method is an accurate interpolation technique that executes the condition that the forecasted value of a point is more influenced by nearer known points than by those which are far away. The general equation for the IDW method is:

$$z_0 = \frac{\sum_{i=1}^{s} z_i \frac{1}{d_i{}^k}}{\sum_{i=1}^{s} \frac{1}{d_i{}^k}} \qquad (1)$$

Where z_0 is the computed value at point 0, z_i is the z value at known point I, d_i is the distance between point I and point 0, s is the number of known points used in interpolation, and k is the specified power.

Spline uses a mathematical function to compute unknown values by minimizing the overall surface curvature [20]. This enables a smooth surface formed exactly accommodating all the points. This interpolation method uses a different form of polynomial interpolation and is best for surfaces that change gradually. It is the best tool for representing gradually varying surfaces with small errors because it can forecast valleys and ridges in the dataset.

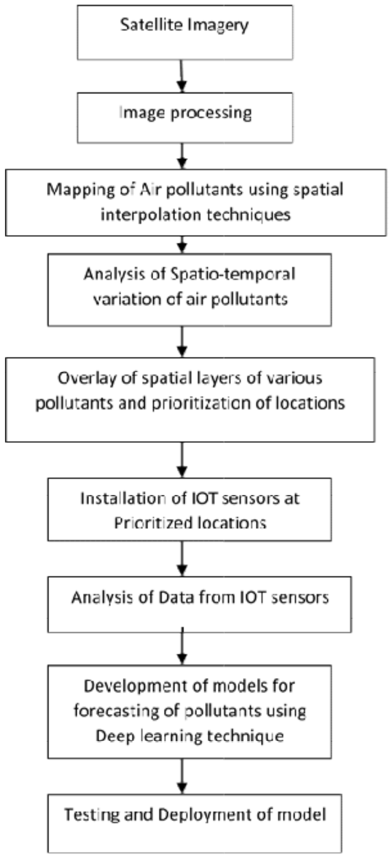

Fig. 1. Methodology Flowchart

Kriging is a spatial interpolation technique that uses a Gaussian method with prior covariance to model the interpolated values [1]. This stochastic approach can be used in a variety of situations, including disaster prevention, geochemistry, and pollution modelling [6].It's based on the idea that the distance between different sample points is spatially associated, indicating surface variation. Random noise is reduced using this form. It calculates the values for all points within a predetermined range. It forecasts values from observed data using a more sophisticated weighted average approach. Mathematically, this function is represented as

$$Z^*(u) - m(u) = \sum_{i=1}^{n(u)} \lambda_i [Z(u_i) - m(u_i)] \text{---- (2)}$$

where u and u_i are location vectors for computation of point and one of the neighbouring data points, indexed by i; n (u) is the total number of data points in the local neighbourhood used for estimation of Z*(u); m(u) and m(ui) are expected values (means) of Z(u) and Z(ui); and λi(u) is the Kriging weight assigned for datum Z(ui) for estimation of location u; the same datum will receive a different weight for different estimation location.

B. Air Pollutant Forecasting

Multidate images are needed for forecasting of the air pollutants. All images are pre-processed using the following steps.

- Acquisition of Images – It involves acquiring and storing of images to database. Further converting it into a variable, creation of load folders containing images into arrays.
- Resizing Images – In order to bridge the mismatch in the requirement of our model and the actual image captured, image resizing is necessary.
- Denoising the images – Unwanted image noises are to be curbed. Smoothening of images is a requirement in order to enhance different scales image structures.
- Segmentation and Morphological changes – Image segmentation may be required for separating the foreground from the background for extracting useful features. Edge Smoothening is required for implementing morphological changes.

After pre-processing the images can be used to simulate a model to forecast the concentration of air using Artificial Intelligence and machine learning techniques. As a sub field of Machine learning, Deep learning has gained tremendous impetus in finding applications in various fields like Natural Language Processing (NLP), healthcare, bioinformatics and exposure science over these years. Its ability of employing numerous distinct deep neural network architectures has made it a popular choice to get state of the art results in varied applications. With a mathematical foundation of algebra and calculus it implements neural networks with more than a single hidden layer of neuron. Deep learning model building will be relying on these 5 important steps as mentioned below:

Defining the architecture: For defining the architecture the nature of problem is to be studied. Due to the nature of the problem studied, Convolution Neural Networks (CNN) happens to be a popular choice to carry out work in image segmentation and image classification on which detailed prediction work can be carried out [18]. The model building structure used for the entire deep learning architecture is Sequential Models, Functional API, or custom architecture which can be defined.

Compilation of the model – To configure the model for the fitting/training process compilation of the model should be done. Few of the critical components of the training procedure get defined for the evaluation procedure during the compilation step. Due to the nature of the problem we will be subjected to losses which are to be defined at this step, we also need to decide on the optimizers and metrics such as accuracy and classification related metrics are to be defined.

Step 3: Fitting the model – To fit the model on the training dataset happens to be an important procedure. It trains the model for a fixed number of epochs (iterations on the dataset).During the entire training procedure the fitting step must be continuously evaluated. It is essential to make sure the model being trained is performing well with improving accuracies as well as a reduction in the overall loss. Over-fitting the model will also be avoided. Tools like Tensor board will be handy for such a procedure.

Step 4: Evaluating and Making Predictions — evaluating the performance of the deep learning model in real scenarios of our application is important and will be carried out. We need to ensure that the predictions made by our model on the test data (split at the beginning of the pre-processing step) are taken into consideration for the purposes of validating the effectiveness of the trained model. Further Random tests will further substantiate the effectiveness on untrained data.

Step 5 Deploy the model – As a final stage of any model construction it has to be deployed. We plan to deploy it as an android and web application building it with either flask, Django, or any other similar framework.

IV. Conclusion

The rapidly increasing concentration of air pollutant levels in urban cities is alarming; hence air pollution monitoring study is imperative. Satellite remote sensing technique enables the quantification of the concentration of air pollutants and locations can be prioritized based on GIS based spatial interpolation techniques. Moreover, having considered the calamitous effects of air pollutants on human health, it is important to forecast air pollutant concentrations by simulation of models. This study was aimed at providing a systematic review of the satellite remote sensing techniques in mapping and prioritizing air pollutant levels and also to predict the concentration of air pollutants. It will help people in the meteorological department and other Government and private agencies to forecast air pollution levels so that further action can be taken accordingly.

References

[1]. A. Satter, “Monitoring Air Pollution Using Satellite Data,” in Proc. *International Conference on Industrial Engineering and Operations Management* Riyadh, Saudi Arabia, 2019.

[2]. T.C. Bailey, and A.C. Gatrell. *Interactive Spatial Data Analysis.* London, UK: Longman. 1995.

[3]. S. Beirle, K. F. Boersma, U. Platt, M. G. Lawrence, and T. Wagner, “Megacity Emissions and Lifetimes of Nitrogen Oxides Probed from Space,” *Science*, vol. 333, pp. 1737-1739, 2011.

[4]. K. F. Boersma, D. J. Jacob, E. J. Bucsela, A. E. Perring, R. Dirksen, R. M. Yantosca, R. J. Park, M. O. Wenig, T. H. Bertram, and R. C. Cohen, “Validation of OMI tropospheric NO2 observations during INTEX-B and application to constrain NOx emissions over the eastern United States and Mexico,” *Atmos. Environ.*, vol. 42, no. 19, pp. 4480-4497, 2008.

[5]. M. Castelli, F. M. Clemente, A. Popovic, Silva, S., Vanneschi, L, A Machine Learning Approach to Predict Air Quality in California, *Hindawi,* Vol. 2020, Article ID 8049504,https://doi.org/10.1155/2020/8049504, 2020.

[6]. R. L. Curier, R. Kranenburg, A. J. S. Segers, R. M. A. Timmermans, and M. Schaap, “Synergistic use of OMI NO2 tropospheric columns and LOTOS–EUROS to evaluate the NOx emission trends across Europe,” *Remote Sensing of Environment*, vol.149, pp. 58-69, 2014.

[7]. J.K. Deters, R. Zalakeviciute, M. Gonzalez, Y. Rybarczyk, “Modeling PM2.5 Urban Pollution Using Machine Learning and Selected Meteorological Parameters,” *Journal of Electrical and Computer Engineering*, Article ID 5106045,https://doi.org/10.1155/2017/5106045, 2017.

[8]. T. H. Do, E. Tsiligianni, X. Qin, J. Hofman, V. P. La Manna, W. Philips, and N. Deligiannis, “Graph-Deep-Learning-Based Inference of Fine-Grained Air Quality from Mobile IoT Sensors.” *IEEE Internet of Things Journal*, vol.7, no. 9, 8943-8955. (2020).

[9]. V.E. Fioletov, C.A. McLinden, N. Krotkov, M.D. Moran, K. Yang, "Estimation of SO2 emissions using OMI retrievals," *Geophys. Res. Lett.* Vol. 38, L21811, 2011.

[10]. D. L. Goldberg, L. N. Lamsal, C. P. Loughner, W. H. Swartz, Z. Lu, and D. G. Streets, "A high-resolution and observationally constrained OMI NO2 satellite retrieval," *Atmospheric Chemistry and Physics*, vol. 17, pp. 11403–11421, 2017.

[11]. J. F. Grajales, and A. Baquero-bernal, "Inference of surface concentrations of nitrogen dioxide (NO2) in Colombia from tropospheric columns of the ozone measurement instrument (OMI)," *Atmósfera,* vol. 27, no. 2, pp. 193–214, 2014.

[12]. H. Greiner, "A survey on univariate data interpolation and approximation by splines of given shape." *Math. Comput. Model.* Vol.15, no. 10: pp. 97–108.doi:10.1016/0895-7177 (91)90094-N. 1991.

[13]. D.A. Griffith, *Advanced Spatial Statistics*. Dordrecht, The Netherlands: Kluwer Academic, 1988.

[14]. J.L. Gunnink, and P.A. Burrough "Interactive spatial analysis of soil attribute patterns using exploratory data analysis (EDA) and GIS." *In Spatial Analytical Perspectives on GIS*, ed. I. Masser and F. Salge,. New York, N Y: Taylor & Francis. pp. 87–99. 1996.

[15]. D. G. Hadjimitsis, K. Themistocleous, and A. Nisantzi, "Air Pollution Monitoring Using Earth Observation & GIS," *Air Pollution.* Rijeka: IntechOpen, 2012.

[16]. D.K. Jha, M. Sabesan, A. Das, N.V. Vinithkumar, and R. Kirubagaran, "Evaluation of interpolation technique for air quality parameters in Port Blair, India." *Univ. J. Environ. Res. Technol*. vol. 1, pp. 301–310. 2011.

[17]. K. Kaya, and Ş. G. Öğüdücü, "Deep Flexible Sequential (DFS) Model for Air Pollution Forecasting." *Scientific Reports,* vol. 10, no. 1, pp. 1-12, 2020.

[18]. X. Li, L. Peng, Y. Hu, "Deep learning architecture for air quality predictions." *Environ Sci Pollut Res* vol. 23, pp. 22408–22417, 2016 https://doi.org/10.1007/s11356-016-7812-9.

[19]. S. Michaelides, D. Paronis, A. Retails, F. Tymvios, "Monitoring and Forecasting Air Pollution Levels by Exploiting Satellite, Ground-Based, and Synoptic Data, Elaborated with Regression Models," Hindawi, vol. 2017, https://doi.org/10.1155/2017/2954010.

[20]. D. Qin, J. Yu, G. Zou, R. Yong, Q. Zhao, and B. Zhang, "A novel combined prediction scheme based on CNN and LSTM for urban PM 2.5 concentration." IEEE Access, vol. 7, pp. 20050-20059, 2019.

[21]. S. Weichenthal, M. Hatzopoulou, and M. Brauer, "A picture tells a thousand exposures: opportunities and challenges of deep learning image analyses in exposure science and environmental epidemiology" Environment international, vol. 122, pp. 3-10, 2019.

[22]. D.W. Wong, L. Yuan, and S.A. Perlin, "Comparison of spatial interpolation methods for the estimation of air quality data." J. Expos. Anal. Environ. Epidemiol. vol.14, no. 5, pp. 404–415. 2004.

[23]. World Health Organization (WHO). Air Quality and Health. Fact Sheet N 313. Updated May 2018.

[24]. L. Zhang, D. J. Jacob, K. F. Boersma, D. A. Jaffe, J. R. Olson, K. W. Bowman, J. R. Worden, A. M. Thompson, M. A. Avery, R. C. Cohen, J. E. Dibb, F. M. Flock, H. E. Fuelberg, L. G. Huey, W. McMillan, H. B. Singh, and A. J. Weinheimer, "Transpacific transport of ozone pollution and the effect of recent Asian emission increases on air quality in North America: an integrated analysis using satellite, aircraft, ozonesonde, and surface observations," Atmos. Chem. Phys., vol. 8, pp. 6117-6136, 2008.

[25]. Y. Zhou, F. J. Chang, L. C. Chang, I. F. Kao, and Y. S. Wang, "Explore a deep learning multi-output neural network for regional multi-step-ahead air quality forecasts", Journal of cleaner production, vol.209, pp. 134-145, 2019.