Research Article

Journal of
Environmental
Nanotechnology

Deep Learning-based Environmental Air Quality Development using Remote Sensing based Stations

M. Mathan Kumar[1*], S. Gayathri Priya[2], S. Sabarunisha Begum[3], Chandrasekar Perumal[4], P. Ramadevi[5] and R. Sudha[6]

[1]Department of Computing Technologies, SRM University, KTR Campus, Chennai, TN, India
[2]Department of Electronics and Communication Engineering, R.M.D. Engineering College, Chennai, TN, India
[3]Department of Biotechnology, P.S.R. Engineering College, Sivakasi, TN, India
[4]Department of Electrical and Electronics Engineering, Vel Tech Rangarajan, Dr.Sagunthala R&D Institute of Science and Technology, TN, India
[5]Department of Electronics and Communication Engineering, University College of Engineering, BIT Campus, Anna University, Tiruchirappalli, TN, India
[6]Department of Information Technology, PSNA College of Engineering and Technology, Dindigul, TN, India
Received: 05.05.2024 Accepted: 06.06.2024 Published: 30.06.2024
*drmathan16@gmail.com

ABSTRACT

Air pollution, comprised of harmful substances suspended in the air, tragically leads to millions of premature deaths on an annual basis. Although ground-based stations offer precise monitoring of air pollution, their effectiveness is confined to specific geographic areas. Conversely, satellite remote sensing technology holds the promise of broadening coverage, yet its primary focus remains on the upper layers of the atmosphere. In pursuit of a comprehensive solution, this study endeavors to revolutionize the representation of surface air quality on a global scale by harnessing the power of satellite data. Through the utilization of advanced techniques such as Generative Adversarial Networks (GANs) and Convolutional Neural Networks (CNNs), it amalgamates data streams from air pollution stations across the globe. Furthermore, socioeconomic and environmental data are seamlessly integrated with satellite images to construct sophisticated multiple models. The results of this innovative approach unveil the superiority of multiple models over their singular counterparts, boasting enhanced accuracy in air quality prediction. The efficient architecture of CNNs combined with the generative capabilities of GANs enables real-time or near-real-time monitoring of air pollution levels. This timely feedback is essential for implementing prompt interventions and mitigating the impact of air pollution on public health and the environment.

Keywords: Generative adversarial networks; Air quality; Satellite remote sensing technology; Convolutional neural networks; Environment prevention.

1. INTRODUCTION

The quality of our environment, particularly the air we breathe, is a critical determinant of public health and well-being. With the rapid advancements in technology, particularly in the realms of deep learning and remote sensing, there exists an unprecedented opportunity to revolutionize the monitoring and management of environmental air quality. Deep learning, a subset of artificial intelligence inspired by the structure and function of the human brain, has emerged as a powerful tool for extracting meaningful insights from complex datasets. By leveraging deep learning algorithms (Zhang *et al.* 2020), we aim to analyze vast amounts of environmental data collected from remote sensing platforms to predict air quality parameters such as pollutant concentrations, aerosol optical depth, and particulate matter levels. The integration of deep learning with remote sensing data holds immense promise for improving the accuracy and timeliness of air quality monitoring and forecasting (Shen *et al.* 2020). By harnessing the wealth of information available from satellites, drones, and other remote sensing platforms, we may learn more about the spatial and temporal dynamics of air pollution and how it affects people and the planet. Ischemic heart disease, stroke, lung cancer, and chronic obstructive pulmonary disease are all made worse by particulate matter, which significantly raises the chance of unexpected myocardial infarction.

However, because of their low spatial resolution, these monitors are often placed in regions that are thought to be potential hotspots for pollution. The fact that there are a number of nations without air quality monitoring stations only makes the problem worse. Another reason to look at other approaches is that the maintenance costs might be deterrent to their broad deployment. To facilitate large-scale observations and reduce spatial distribution uncertainty, remote sensing technology has become an attractive choice because of the vast areas it covers. Nevertheless, the ability of remote sensing to detect air pollution is mostly limited to the Earth's upper atmosphere, where measurements are highly sensitive. However, these findings may still be used to depict changes in the distribution of air pollution on Earth's surface (Filonchyk *et al.* 2021; Sakti *et al.* 2023) For each kind of pollutant (PM, NO_x, and SO_2) measured in 2007 in Hyderabad, India, the API from the study by Mozumder *et al.* (2013) is defined as an average comparison value between the present and a standard. From July 2018 through July 2021 in Poland, Grzybowski *et al.* (2021) assessed NO_2 pollution concentrations in the field by merging data from the Sentinel-5P satellite and data on driving conditions. They used several linear regression and machine learning algorithms. Using data from Sentinel-5P, Sentinel-2, and local field stations, Rowley *et al.* (2023) estimated NO_2, O_3, and PM10 air pollution in the UK and Ireland using a multimodal-artificial intelligence architecture. The difficulty in constructing a globally applicable model for air pollution assessment stems from the need to incorporate the specific climatic, geographical and socioeconomic factors of each site.

1.1 Contribution of the Research

- By applying deep learning methodologies to environmental data from remote sensing sources, this research contributes to the ongoing evolution of machine learning techniques for environmental monitoring. Specifically, we explore novel architectures and algorithms tailored to the unique challenges of air quality prediction, thereby expanding the range of tools available to researchers and practitioners in this domain.
- Traditional air quality monitoring methods are often limited by spatial coverage and temporal resolution. By leveraging data from remote sensing platforms, this research enables the generation of high-resolution, real-time air quality maps over large geographic areas. This enhanced spatial and temporal resolution provides policymakers, urban planners, and public health officials with valuable insights into localized pollution hotspots and trends, facilitating more targeted interventions and mitigation strategies.
- The integration of deep learning with remote sensing data enables more accurate and reliable predictions of air quality parameters such as pollutant concentrations and particulate matter levels. By harnessing the power of machine learning algorithms to analyze complex environmental datasets, this research enhances our ability to forecast air quality conditions with greater precision, thereby supporting informed decision-making and risk assessment.
- The findings of this research have important implications for air quality management and policy development. By providing timely and accurate information on air pollution levels, this research empowers the concerned to implement targeted interventions and regulatory measures aimed at reducing emissions and protecting public health. Additionally, by raising awareness of the link between environmental factors and air quality, this research contributes to the broader discourse on sustainability and environmental monitoring.
- The methodologies and techniques developed in this research are designed to be scalable and replicable, allowing for widespread adoption and application across diverse geographical regions and environmental contexts. By establishing a framework for integrating deep learning with remote sensing data for air quality monitoring, this research lays the groundwork for future studies and initiatives aimed at addressing environmental challenges on a global scale.

2. PROPOSED METHODOLOGY

The proposed methodology for this research involves several key steps aimed at leveraging deep learning techniques in conjunction with data from remote sensing-based stations for environmental air quality development. Firstly, we will acquire and preprocess environmental data from remote sensing platforms, including satellite images and ground-based sensor networks. This data will encompass a range of environmental variables, including atmospheric composition, meteorological parameters, and land use characteristics, to provide a comprehensive view of air quality dynamics. Next, we will design and implement deep learning architectures tailored to the task of air quality prediction. This will involve exploring various neural network architectures, such as CNN and GAN, and optimizing hyper parameters to maximize model performance. Additionally, transfer learning techniques may be employed to leverage pre-trained models and accelerate training on limited datasets. Once the deep learning models are trained, we will evaluate their performance using a combination of quantitative metrics and qualitative analysis. This will involve comparing model predictions against ground truth air quality measurements collected from established monitoring stations and assessing the models' ability to generalize to unseen data. To validate the robustness and generalizability of the proposed methodology, we will conduct case studies in diverse geographical regions with varying environmental characteristics. This will allow us to assess the transferability of the models across different spatial and temporal scales and identify potential sources of bias or uncertainty.

Fig.1 shows a flowchart for the suggested model's early warning method. The technique begins with data cleansing, which involves removing undesired and abnormal data. The early warning system operates in three modules: pre-processing, forecasting, and evaluation. Finally, we will interpret our research findings in light of existing literature and address the implications for air quality management and policy. This will entail identifying critical insights acquired from deep learning models and outlining opportunities for future research and development. To summarise, the suggested methodology combines cutting-edge techniques from deep learning and remote sensing to create a revolutionary approach to environmental air quality enhancement. We hope to increase our understanding of air quality dynamics and promote evidence-based decision-making in environmental management and public health by using the amount of data accessible from remote sensing systems and harnessing the power of machine learning algorithms.

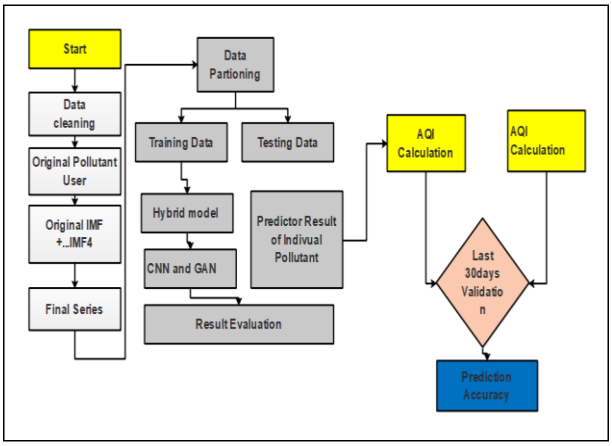

Fig. 1: Flowchart of Proposed Model

2.1 Data Collection

From 1 March 2019 to 1 March 2021, the information containing air pollutant concentrations and meteorological conditions in India was received from the Central Pollution Control Board (CPCB) in Delhi. Data was gathered from 27 monitoring stations for seven different air pollutants: SO2, NO, NO_2, O_3, PM10, PM2.5, CO, and six meteorological parameters: wind speed (WS), wind direction (WD), solar radiation (SR), pressure (BP), atmospheric temperature (AT), and relative humidity (RH).

2.2 Data Pre-processing

Data for many days was missing due to a combination of causes, including instrumental mistakes, measurement problems, data transmission difficulties, and others. Though there are forty monitoring stations in Delhi, only those with complete data were used for evaluation because of missing values in contaminants and weather.

2.3 Description of Monitoring Sites

The National Capital Region (NCR) is home to many air quality monitoring stations. The CPCB, DPCC, and SAFAR from IITM, Pune are in charge of the monitoring. Sections such as Sarojini Nagar, Chandni Chowk, Mayapuri Industrial Area, Pitampura, Shahdara, Shahzada Bagh, Nizamuddin, Janakpuri, Siri Fort, and ITO are monitored by the NAMP of the CPCB. Anand Vihar, Civil Lines, DCE, Dilshad Garden, Dwarka, IGI Airport, ITO, Mandir Marg, Punjabi Bagh, R.K. Puram, and Shadipur are among the eleven regions that CAAQM keeps an eye on, while DPCC watches over six different spots, including Civil Lines, Punjabi Bagh, Mandir Marg, Anand Vihar ISBT, IGI Airport, and R.K. Puram. The data from the various monitoring stations is kept by DPCC and IMD for Pusa, CPCB for NSIT Dwarka and Siri Fort, and DPCC for the rest of the stations. In addition to the CPCB and DPCC, SAFAR maintains eight monitoring stations all throughout Delhi to track the air quality in real-time. One such thing that uses these stations' data is the national air quality index.

2.4 Mathematical Modelling

One way to model the player population is with a matrix, where each row characterizes a player besides each column characterizes their various qualities. There are exactly as many columns in this matrix as there are problem variables and the values that have been proposed for them. Using (1), we can specify the players' matrix.

$$X = \begin{bmatrix} X_1 & x_1^1 & \cdots & x_1^d & \cdots & x_1^m \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ X_i & x_i^1 & \cdots & x_i^d & \cdots & x_i^m \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ X_N & x_N^1 & \cdots & x_N^d & \cdots & x_N^m \end{bmatrix} \tag{1}$$

Here X is matrix, x_i^d is the d^{th} breadth of i^{th} player, m is the sum of variables, besides N is the sum of players. By including X_i in function, valuable insights are derived, as shown in (2) to (7).

$$F_{best} = min \tag{2}$$

$$X_{best} = X(location of min(fit), 1{:}m) \tag{3}$$

$$F_{worst} = max \tag{4}$$

$$X(location of max(fit), 1{:}m) \tag{5}$$

$$F^n = \frac{fit - F_{worst}}{\sum_{j=1}^{N}(fit_j - F_{worst})} \tag{6}$$

$$P_i = \frac{F_i^n}{max(F^n)} \tag{7}$$

Here, F_{best} characterizes the best fitness function value, X_{best} is the finest variables' values, F_{worst} is the worst fitness purpose value, X_{worst} is the worst variables' standards, F^n is the regularize functions, and P_i is the likelihood purpose of i^{th} player.

3. RESULTS AND DISCUSSION

The training process utilized a 12-gigabyte NVIDIA Tesla K80 GPU equipped with GDDR5 VRAM, providing 11.439 gigabytes of usable memory. Tensor Flow 2.0, integrated with the Keras API, served as the primary framework for training, evaluating, and predicting across all models in this study. Here are the expansions for the given acronyms:

- DPCC: Delhi Pollution Control Committee
- SAFAR: System of Air Quality and Weather Forecasting and Research
- NAMP: National Air Quality Monitoring Programme
- DCE: Delhi College of Engineering (now known as Delhi Technological University)
- ITO: Income Tax Office (often refers to the ITO area in Delhi, where several government offices are located)
- CAAQM: Continuous Ambient Air Quality Monitoring
- ISBI: Institute of Systems Biology India
- IMD: India Meteorological Department
- NSIT: NetajiSubhas Institute of Technology (now known as NetajiSubhas University of Technology)

3.1 Evaluation Module

After the forecasting module, the results were evaluated using two criteria: Standard evaluation parameters and AQI assessment. RMSE, MSE, MAPE, MDA, and MDAPE were used to evaluate the performance of the proposed model.

3.1.1 RMSE

A commonly used statistic for gauging the discrepancy between a model's projected values and the actual values is the Root-Mean-Square Error. The root-mean-squared error (RMSE) may be determined by calculating the following: the mean of the residuals, the square root of the mean, the norm of the residuals for each data point, and the residuals themselves, which represent the difference between the forecast and reality.

3.1.2. MAE

The mean absolute error (MAE) is a measure of the size of errors for a collection of predictions and observations based on the mean of the absolute errors for the group. It is also known as the L1 loss function.

3.1.3. MAPE

It is often referred to as Mean Absolute Percentage Error (MAPE). In statistics, it is a measure of the prediction accuracy of a forecasting approach. First, the absolute difference between the Actual Value (At) and the Predicted value (Pt) are determined. The mean function is then applied to the result to obtain the MAPE value. The lower the MAPE, the better the model's fit.

3.1.4. MDA

Mean directional accuracy, also referred to as mean direction accuracy, is a statistical measure of the prediction accuracy of a forecasting approach. The predicted direction (upward or downward) and the actual obtained direction are compared.

3.2 Standard Evaluation Parameter Criteria

Root-Mean-Square Error (RMSE) is a commonly used statistic to quantify the discrepancy between the values that a model predicts and the actual values. The recommended hybrid model has reduced the RMSE value for all air pollutants, with CO having a value of 0.0527, NO of 0.0775, NO_2 of 0.0441, SO_2 of 0.0298, O_3 of 0.0622, PM2.5 of 0.0594, and PM10 of 0.0743. To calculate RMSE, one must know the residual (the difference between prediction and reality) for each data point, as well as the norm of residual, mean of residuals, and square root of the mean.

Tables 1 to 4 represent that the validation of different measure performance with different classifier models and proposed model as RMSE, MAE, MAPE and MDA. In the analysis of different gas as CO, NO, NO_2, SO_2, O_3 and PM10. In MDA calculation of O_3 pollutant, the LSTM model attained as 91.27 and LSTM model attained as 93.12 and the proposed model as 93.12 correspondingly. Then the PM10pollutant, the LSTM model attained as 95.66 and GRU model attained as 95.66 and the proposed model as 97.66 correspondingly.

Table 1. Validation Criteria: RMSE

Validation Criteria	Pollutant	LSTM	GRU	Proposed
RMSE	CO	87.3	87.3	87.47
	NO	88.22	88.29	88.22
	NO_2	89.86	89.24	89.86
	SO_2	91.25	91.27	91.28
	O_3	93.12	93.12	93.12
	PM10	95.66	95.66	97.66

Table 2. Validation Criteria: MAE

Validation Criteria	Pollutant	LSTM	GRU	Proposed
MAE	CO	87.3	87.3	87.30
	NO	88.58	88.29	88.22
	NO_2	89.86	89.86	89.81
	SO_2	91.25	91.27	91.28
	O_3	93.12	93.12	93.12
	PM10	95.66	95.66	97.66

Table 3. Validation Criteria: MAPE

Validation Criteria	Pollutant	LSTM	GRU	Proposed
MAPE	CO	87.3	87.3	87.3
	NO	88.15	88.29	88.25
	NO_2	89.86	88.29	89.22
	SO_2	91.31	89.86	88.29
	O_3	93.12	91.27	89.86
	PM10	95.66	95.66	97.66

Table 4. Validation Criteria: MDA

Validation Criteria	Pollutant	LSTM	GRU	Proposed
MDA	CO	87.3	87.3	87.3
	NO	88.22	88.29	88.29
	NO_2	88.29	89.86	89.86
	SO_2	89.86	91.27	91.27
	O_3	91.27	93.12	93.12
	PM10	95.66	95.66	97.66

4. CONCLUSION

The integration of deep learning with remote sensing data for environmental air quality monitoring represents a significant leap forward in our ability to understand, predict, and manage air pollution. Through the development and application of advanced machine learning algorithms, this research has demonstrated the potential to enhance monitoring precision, provide timely alerts, and inform evidence-based policymaking in the realm of environmental health. By leveraging the vast amount of data available from remote sensing platforms, we have been able to gain deeper insights into the complex dynamics of air quality, uncovering patterns and trends that were previously obscured by limitations in spatial and temporal resolution. These insights have not only advanced our scientific understanding of air pollution but also empowered policymakers, urban planners, and public health officials to implement targeted interventions aimed at reducing emissions and protecting public health. As we look to the future, continued research and innovation in the fields of deep learning and remote sensing will be essential for addressing emerging environmental challenges and safeguarding the health and well-being of current and future generations. By building upon the foundations laid by this research, we can work towards a more sustainable future, where clean air is a fundamental right enjoyed by all. In conclusion, the findings of this research underscore the transformative potential of harnessing cutting-edge technologies for environmental monitoring and management. Through collaboration across disciplines and sectors, we can harness the power of data-driven insights to tackle the pressing environmental issues of our time.

FUNDING

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

CONFLICTS OF INTEREST

The authors declare that there is no conflict of interest.

COPYRIGHT

REFERENCES

Filonchyk, M., Hurynovich, V., Yan, H., Impact of Covid-19 lockdown on air quality in the Poland, Eastern Europe, *Environ. Res.* 198, 110454 (2021). https://doi.org/10.1016/J.ENVRES.2020.110454

Grzybowski, P. T., Markowicz, K. M., Musiał, J. P., Reduction of Air Pollution in Poland in Spring 2020 during the Lockdown Caused by the COVID-19 Pandemic, *Remote Sens. 2021, Vol. 13, Page 3784* 13(18), 3784 (2021). https://doi.org/10.3390/RS13183784

Mozumder, C., Reddy, K. V., Pratap, D., Air Pollution Modeling from Remotely Sensed Data Using Regression Techniques, *J. Indian Soc. Remote Sens.* 41(2), 269–277 (2013). https://doi.org/10.1007/S12524-012-0235-2

Rowley, A., Karakuş, O., Predicting air quality via multimodal AI and satellite imagery, *Remote Sens. Environ.* 293, 113609 (2023). https://doi.org/10.1016/J.RSE.2023.113609

Sakti, D. H., Cornish, E. E., Fraser, C. L., Nash, B. M., Sandercoe, T. M., Jones, M. M., Rowe, N. A., Jamieson, R. V., Johnson, A. M., Grigg, J. R., Early recognition of CLN3 disease facilitated by visual electrophysiology and multimodal imaging, *Doc. Ophthalmol.* 146(3), 241–256 (2023). https://doi.org/10.1007/S10633-023-09930-1/FIGURES/4

Shen, H., Jiang, Y., Li, T., Cheng, Q., Zeng, C., Zhang, L., Deep learning-based air temperature mapping by fusing remote sensing, station, simulation and socioeconomic data, *Remote Sens. Environ.* 240, 111692 (2020). https://doi.org/10.1016/J.RSE.2020.111692

Zhang, Q., Fu, F., Tian, R., A deep learning and image-based model for air quality estimation, *Sci. Total Environ.* 724, 138178 (2020). https://doi.org/10.1016/J.SCITOTENV.2020.138178