ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20231016114033

Original scientific paper
Received: 2023-10-16, Accepted: 2023-11-17

Forecasting Daily Air Quality Index and Early Warning System for Estimating Ambient Air Pollution on Road Networks Using Gaussian Dispersion Model with Deep Learning Algorithm

Asha Unnikrishnan*, S. Rajeswari

Abstract: The rapid growth of the vehicle population is a major factor in heavy air pollution and public health issues. Traffic-related air pollutants (TRAPs) on roads are often much higher than ambient values, leading to high exposure levels in vehicles. This research proposes a hybrid forecasting model for early detection and early warning systems (EWS) of road networks during real-world travels. Data is collected from Kannur, Calicut, Palakkad, and Coimbatore using real-time sensors, including surrounding discussion information, activity information, vehicle speed, and stopping events. The study predicts ambient air quality (AAQ) levels on the road network using the Gaussian Dispersion model (GDM) and measures the risk sensitivity of PM10 and PM2.5 in selected regions. This helps formulate powerful prevention strategies and prevent negative health impacts. The air pollution module for predicting concentration has an innovative hybridization model that combines an improved cuckoo search (CS) and differential evolution (DE) algorithm with a stacked LSTM model to increase forecasting accuracy of six major environmental pollution levels. This model predicts the AAQ level and is effective and robust for warning one day before the pollutant event occurs based on the risk level of an ambient air pollutant from the RN.

Keywords: Air Quality Index (AQI); Ambient Air Pollution (AAP); EWS; GDM; Hybrid Forecasting Method; LSTM; Road Networks

1 INTRODUCTION

Atmospheric oxidation capacity (AOC) is the main factor promoting the development of multidimensional air pollution in the troposphere and the near-surface atmosphere, it regulates both the pace at which trace gases are removed and the rate at which secondary pollutants are produced [1]. Governmental agencies use the AQI to estimate air pollution levels and educate the public. The growth in AQI is anticipated to have major negative health implications on a sizeable section of the population [2]. A pollution concentration and an air monitor during the average period that is predetermined are necessary for the AQI measurement [3]. An EWS uses integrated communication technologies as a climate change adaptation strategy to assist communities with potentially dangerous climate-related events. Long-term sustainability is supported by the successful EWS by preserving people, jobs, land, and infrastructure [4]. Initiator monitoring poses major hazards, such as those brought on by the use of CBRN (biological, radiological, chemical, and nuclear) agents, which are crucial given the increasing threat of war and terrorism [5]. There is a need for safety systems that can identify these dangerous substances, continuously monitor the air, and protect people from them [6]. It has become vital to install stationary pollution and radiation identification sensors in public utility facilities, vehicles, and buildings so that continuous air quality monitoring and protection is carried out at significant places [7].

Pollutant levels such as $PM_{2.5}$, NO_2, and SO_2 can be determined using data obtained from various air quality monitoring stations [8]. Some gases, such as CO_2 and CO can even be found using portable sensors. An AQI and a category, such as good, average, and unhealthy, are used to quantify air quality [9]. The effect of numerous factors, including traffic flow, fast urbanisation, land use, massive industry, and population density causes large location- and time-specific variations in urban air quality [10]. Every country's economic development depends on a large part of its RN. To provide reliable connections between the various portions of geographical territory, it is crucial to foresee a deliberate and ongoing extension and adequate maintenance of these networks [11]. Air quality is negatively impacted by traffic-related concerns such as outdoor parking and the density of the system of roads [12]. A Geographic Information System (GIS) that incorporates sub-models enables the use of geographical coordinates to represent the layout of urban areas, transportation networks, and the dispersion of pollutants in the atmosphere [13]. Monitoring the polluting gases regularly is vital to control air pollution. In the literature, several sensor network-based air contamination supervision systems have been developed, implemented, and evaluated [14]. The methodology employed the evaluation attributes, the testing location, the pollutants assessed using sensors, the system performance, the communication device used, and the microcontroller used are all considered when compared to the pollution monitoring systems [15]. To evaluate AAP on RN, the research will analyse the index of air quality and EWS using a GDM and deep learning algorithm. The remainder of the report is organised as follows: Section 2 describes the research's literature review, and Section 3 describes the problem definition and research's rationale. The proposed approach is shown in section 4, the experimentation and findings are shown in section 5, and the research's conclusion is explained in section 6.

2 LITERATURE SURVEY

The spatial qualities of each type of road and the average annual PM_{10} concentration relationship were investigated by Sohrab et al. [16]. The study's findings are utilized to reduce conflicts between the environment and transport that affect air quality in urban, urban-rural fringe, and rural (agricultural) environments. Accordingly, Song et al. [17] evaluated the relative value of features, the meteorological set of characteristics is the most relevant, followed by the land use features. Using a small number of air pollution monitoring stations, the suggested Multi-AP approach might be used to calculate air pollution exposure in a city. Galán-Madruga et al. [18] investigated a method for estimating benzene air concentrations from the associated factors. The calculated and the regional distribution of the most typical benzene levels were identical. Finally, a synthetic neural network located the AQMN's fixed benzene monitoring locations that were the most representative. To improve human health, an IoT-qualified Environmental Toxicology for Air Pollution Monitoring expending the Artificial Intelligence method (ETAPM-AIT) is developed by Asha et al. [19]. The experimental results demonstrate how well the suggested ETAPM-AIT model outperforms more recent methods. A pattern of inexpensive IoT systems for tracking AQI and traffic flow was presented by Martn-Baos et al. [20]. The suggested architecture is to be validated and enabled with a range of climatic and traffic circumstances, which is the experimentation with data from many cities reflecting various scenarios.

Modern high-resolution European emission inventories for air quality models were studied by Kuenen et al. [21]. Emissions from shipping (both on land and at sea) are estimated using the results of an emission model like separate shipping, which based emissions on real ship movement data and the agricultural waste burning emissions on satellite observations. To forecast short-term local NO_2 and O_3 concentrations, Res-GCN-BiLSTM, a new deep learning (DL) based hybrid model that combines the bidirectional long short-term memory (BiLSTM) and residual neural network (ResNet) was investigated by Wu et al. [22] and graph convolutional network (GCN). The Res-GCN-BiLSTM model performed better than the best-performing baseline model in terms of pollutant adaptation and forecast accuracy, with mean absolute error improvements for NO_2 and O_3 of almost 11% and 17%, respectively. Moses et al. [23] assessed the cadmium and allowed them to monitor the AAQ. According to the study, none of the dust that was spread across the exposed vegetable leaves in each market had any cadmium. To demonstrate the interrelated issues of emissions, behaviour, and the anticipated exposure to PM10 for groups of drivers and underground commuters in Seoul CBD by using an agent-based simulator for traffic Shin et al. [24]. The air quality measurements showed that background PM10 was much lower than roadside PM10 by about 25-30%. For the West Midlands region, Zhong et al. [25] simulated the street-scale resolution air quality. An effective tool for assessing prospective regional mitigation and local air pollution measures is the coupled air quality modelling system for WM. Using the GDM and a DL algorithm, this research study presents an AQI and EWS for evaluating AAP on RNs. Data mining constitutes a stated removal of hidden information about predicting from large files. From the large database, the necessary information is fetched using the data mining process. Here, Neelaveni et al. [26] discussed how data mining techniques are applied in the agriculture field.

3 RESEARCH PROBLEM DEFINITION AND MOTIVATION

A series of environmental issues, including air pollution, deforestation, solid waste management, the release of toxic materials, and many more, have accumulated further consideration than ever before as a finding of urban areas' continuous expansion and population growth. There is an essential for fast information regarding changes in the level of pollution because the issue of air effluence in cities has gotten so severe. One of the causes of air contamination is vehicle exhaust emission during transportation activities. Congestion may be caused by an increase in motor vehicle traffic, a reduction in road space, and activity along the roadside. The primary contributors to air pollution are fossil fuel oil for motor vehicle exhaust emissions. Nitrogen oxide (NOx) in the forms of nitrogen dioxide (NO_2) and nitric oxide (NO), carbon monoxide (CO), carbon dioxide (CO_2), particulate matter (PM_{10}), sulphur oxides (SOX) in the forms of sulphur dioxide (SO_2) and sulphur trioxide (SO_3), and hydrocarbons (HC) are among the pollutants that are released by motor vehicles. These vehicles emit toxic pollutants that, if exposed continuously and over an extended period, have a detrimental effect on human health. Therefore, the goal of this research is to determine how well air pollution levels in the environment may be calculated using the infrastructure's road structural characteristics.

In India, the revelation of ambient particulate thing is the main environmental risk factor. However, Indian cities rank high in the world for PM2.5 air pollution, which has serious negative effects on health. Numerous monitoring locations all over India record continual transgressions of NAAQS breaches and high concentrations (CPCB, 2016). High ambient PM2.5 concentrations have been linked to anthropogenic activities such as fossil fuels burned for electricity generation, industrial processes, road transportation, and traditional residential cooking in India. Without adequate controls, it is conceivable to anticipate additional decays in air superiority due to rising levels of polluting economic activity. Global history shows that clean air may be attained without compromising social and economic progress, though. Numerous research on the impact of traffic-related motor vehicle emissions on air pollution has been done. Through vehicle consumption emission and traffic simulation calculation using Matlab software, the research's goal was to examine the execution of the selected area's RN and vehicle exhaust emissions.

4 PROPOSED RESEARCH METHODOLOGY

In our novel approach for forecasting daily Air Quality Index (AQI) and implementing an Early Warning System (EWS) to estimate ambient air pollution on road networks, we propose a hybrid methodology that combines the Gaussian Dispersion Model with a state-of-the-art Deep Learning Algorithm. Firstly, we leverage the spatial dispersion patterns obtained from the Gaussian model to enhance the understanding of pollutant behaviour across road networks. Subsequently, we integrate a deep learning algorithm, such as a Convolutional Neural Network (CNN), to capture complex temporal and spatial dependencies in the air quality data, ensuring accurate and robust predictions. Our methodology not only provides precise daily AQI forecasts but also establishes a proactive EWS for timely pollution alerts, facilitating effective mitigation strategies.

Monitoring and lowering air pollution begins with the evaluation of air quality. The air supply's suitability for a particular application depends on its properties. Criterion air pollutants are one of the prevalent air pollutants in India. Traffic is extremely diverse on Indian roadways in general and Indian urban highways in particular, with a broad range of static and dynamic characteristics present. Due to the heterogeneity, lane discipline is difficult to enforce, and the vehicles occupy positions on any area of the road depending on available space. Therefore, a suitable modelling method must be created to recreate the traffic flow conditions that have been specified. The complete road space is regarded as a single unit in the proposed simulation technique, without any designated traffic lanes. The vehicles will be visualised as rectangular blocks with dimensions that occupy a predetermined portion of the road space. The coordinates of an origin will be used to indicate the vehicle's position. The interval scanning approach with fixed addition time improvement is the foundation of the simulation model. The length and width of the road stretch can be changed for simulation purposes according to user specifications.

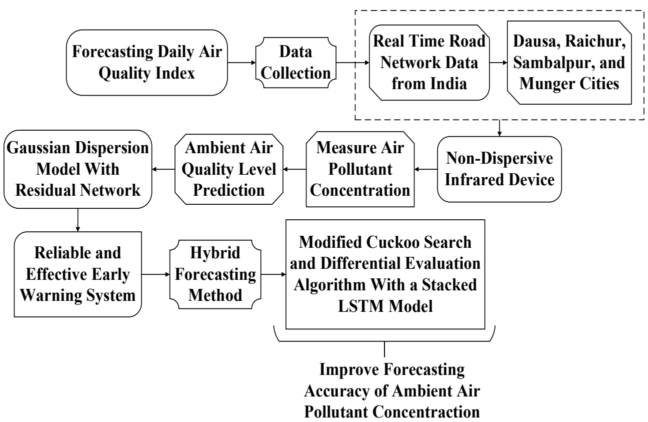


Figure 1 Block Diagram for the Proposed Work

Fig. 1 depicts the flow diagram of the research work. Its accurate real-time air pollution predictions are essential for providing early particle matter warnings. This study determines to forecast that particulate matter concentrations will change over time. A high-volume air sampler (HVAS) device was used to test the concentration of PM10 pollution, a non-dispersive infrared device was used to measure the concentration of CO pollutants, and an impinger tube was used to measure SO_2, O_3, and NO_2 pollutants. DL is a leading statistical technique, which is crucial for anticipating outdoor air quality. The development of prediction models using DL is further guaranteed by the fact that air pollutant concentrations are typically connected with other characteristics, such as SO_2, O_3, CO and NO_2, and these parameters can be detected by real-time sensors. To construct a three-level architecture, this study proposed an end-to-end prediction technique that combines ResNet features with the stacked LSTM variants. The objective is to fully materialise a spatiotemporal association characteristic for PM concentration estimation.

(a) Data Collection: The primary data of this study were obtained directly from the research site, while secondary data were obtained from other sources. The search locations are Kannur, Calicut, Palakkad and Coimbatore Road. The main data used in this study are traffic data such as vehicle speed, traffic volume and number of parking times obtained from a survey at the research site, air pollution data obtained by direct on-site measurements using sensors in collaboration with the Transport Road Research Laboratory and on-road data geometry, obtained by direct measurement at the study site. Actual environmental data in ambient air was collected for three months, from October 2022 to December 2022, as part of the main data collection of the first phase. The weather was sunny with an air temperature of 32 °C at the time the survey collected data.

The surveyors on each road section on the Kannur, Calicut, Palakkad, and Coimbatore RN recorded traffic volume, vehicle speed, and parking events. Traffic data collection also involved traffic volume, vehicle speed, and parking data. Traffic volume data were recorded hourly for each road segment between the hours of 8:00 AM to 10:00 PM. By sampling 30 variable types of vehicles on the selected RN, a speed gun was used to measure the data on vehicle speed. As a result, parking data were acquired by collecting parking events after the entry and exit times of the vehicles at the parking spot in the selected region. The crossroads in Kannur, Calicut, Palakkad, and Coimbatore that had the most traffic and the highest levels of vehicle exhaust emissions were where the ambient air pollutant concentration was measured. Data collection on pollutant content in ambient air took place from 8:00 AM to 10:00 PM for a year, from June 30, 2023, to June 30, 2023. The background for the RN in data processing will be provided by aerial images obtained with Google Earth software as secondary data, respectively.

(b) Ambient Air Quality Level Prediction: When assessing the air quality, the AAQ is crucial. The main air contaminants that, if present at harmful levels, could potentially impair human health are typically included in an air pollution index system. Instead of using the actual concentration of air pollution, an AAQ system is typically constructed in easily understandable ranges of values as a means of reporting the air quality or level of air pollution. To extract the highest level of abstraction from the data, the residual network and GDM are introduced in this study.

(i) GDM: The complex system of different equations that describes the pollutant concentration as a finding of dispersion processes is approximated analytically by the Gaussian plume:

$$\hat{C}(x,y,z)=\frac{Q}{2\pi\sigma_y\sigma_{z\bar{u}}}\exp\left(\frac{-y^2}{2\sigma_y^2}\right)\left[exp\left(-\frac{(z-H)^2}{2\sigma_z^2}\right)+exp\left(-\frac{(z+H)^2}{2\sigma_z^2}\right)\right] \tag{1}$$

Where $\hat{C}$ represents the estimated concentration at a specific location, Q defines the rate of emission of pollutant from a point source at the origin of the coordinate system x, y and z are the crosswind, downwind, and vertical distances from the source. $\hat{u}$ defines the time-averaged fast wind at the effective release height H along the x-axis, whereas σ_Y and σ_Z stand for the crosswind standard deviations. Eq. (2) specifies the effective height of release.

$$H = H_s + \Delta h \tag{2}$$

Where, H_s defines the point of emission's height and Δh is the ascent of the plume. The Gaussian plume model (GPM) is predicated on several hypotheses, such as a uniformly turbulent wind field, steady-state airflow, a steady-state point source of pollution, and total reflection of the pollutant from the ground. It serves as the foundation for several widely used mechanistic air pollution models.

The GPM was parameterized to allow for utilisation as a non-linear regression model. Only vigilant (high-end) estimates are available from regulatory organisations, and the emission rates in this study region are not available with absolute certainty. The actual emission rates are estimated to be αQ where, $0 \leq \alpha \leq 1$ such that the ground-level ($z = 0$) concentration can be stated as follows, presuming that the relative errors in the emission rates are comparable for all sources.

$$\hat{C} = \frac{\alpha Q}{\pi \sigma_y \sigma_{z\bar{u}}} exp\left(\frac{-y^2}{2\sigma_y^2}\right) \exp\left(\frac{H^2}{2\sigma_z^2}\right) \tag{3}$$

There are numerous methods for parametrizing the dispersion standard deviations. Power-law relationships with the downwind distance from the source were selected due to their widespread use, mathematical simplicity, and relatively low number of required parameters.

$$\sigma_y = \frac{x^{p_7}}{\sqrt{2p_6}}, \sigma_z = \frac{x^{p_9}}{\sqrt{2p_8}} \tag{4}$$

Where the optimization parameters are denoted as p_{6-9} . In this concept, emissions from traffic are portrayed as local sources. In the Gaussian formulation, this offset is designated as p_4 . Additionally, it is believed that some unidentified factor p_2 converts traffic attributes to emission rates homogeneously (in space but not in time). The full Gaussian function estimates the concentration at the ith grid cell as described below, taking into consideration all the point and area sources in the study region.

$$\hat{C}_i = p_1 + p_2 \sum_{j=1}^{M} T_j \cdot g_t(x_{ij}, y_{ij}, \hat{u}) + p_3 \sum_{j=1}^{K} Q_j \cdot g_q(x_{ij}, y_{ij}, H_{j,} \hat{u}) \tag{5}$$

Where, p_1 denotes a spatially uniform baseline concentration, T_j denotes the amount of traffic in the jth grid cell, M denotes the number of grid cells in the study area, K denotes the number of point sources in the study area, Q_J denotes the rate of emission from the jth industrial stack, and the following definitions of g_q and g_t are given.

$$g_q(x, y, H, u) = \frac{1}{u \cdot x^{p7+p9}} exp\left(-\frac{p6y^2}{x^{2p7}} - \frac{p8y^2}{x^{2p9}} - p_5 x\right) \tag{6}$$

$$g_t(x, y, u) = \frac{1}{u \cdot (x+p_4)^{p7+p9}} exp - \frac{p_6 y^2}{(x+p_4)^2 p_7} - p_5 x \tag{7}$$

Where, p_{7-9} is utilised to model the dispersion standard deviation and p_5 is the pollutant's first-order removal coefficient. In actuality, atmospheric chemical interactions may be much complex than first-order removal, that as predicted by p_5. However, information like the concentration of hydroxyl radicals, which is frequently unavailable, is needed for an accurate description of these interactions.

(ii) Residual Network: The residual network prevents the disappearance or explosion of gradients by acting as a shallow exponential set. By including shortcut connections, ResNet can be optimised more easily. A few layers that include shortcut connections are referred to as residual blocks. The neural network's conventional input and output are x and $F(x)$, and its training target is $F(x) = H(x)$. In this ResNet, x serves as the neural network's input, $H(x) + x$ serves as its output, and $H(x) = F(x) - x.H(x)$ serves as its training objective. This refers to $H(x)$ as residual. $H(x)$ is simpler to train than $F(x)$. Assuming the size of the input pixel matrix is 32×32, this ResNet unit.

- The convolutional kernel size is 1×1 and there are 64 convolutional kernels in the first layer. Batch Normalisation (BN) and the activation function ReLU() are then applied.
- The second layer's convolutional kernel is 3×3 in size, there are 64 of them, and the subsequent procedure is the same as it was for the first layer.
- There are 256 convolutional kernels in the third layer, with a convolutional kernel size of 11. There is only one 33 convolutional kernel in the shortcut connection. The residual convolution and BN are carried out following the third convolution.
- Finally, a full connection layer is employed to obtain the 32×32×256-sized ResNet output.

Accordingly, at daily time intervals, this model extracts and records spatiotemporal characteristics. A hybrid forecasting model, which is detailed in the following part, is used to input these extracted features into the EWS.

(c) Hybrid Forecasting Method for Early Prediction and Early-Warning System: The new early-warning system for air quality was examined, and its design was based on a hybrid forecasting technique. To generate a precise operating strategy to lessen the negative effects when air pollution events occur, this system is made to obtain correct air quality information in advance.

(i) Modified CS and DE Algorithm: To increase the estimation accuracy, prevent local optima, and optimize the initial weights and thresholds of the stacked LSTM model, a hybrid modified optimization technique between CS and DE is proposed in this part. Here, air pollution concentrations are predicted using a stacked LSTM model. The initial weights and thresholds of the stacked LSTM algorithm were optimized using a CS and DE combined update optimization method to increase convergence stability. The predictive ability was tested using top six ambient air pollutant data compiled in India. Based on the predicted air quality, it successfully operated the early warning system.

Cuckoo Search: With the help of the transformation parameter p_a, the CSA successfully integrates the local stochastic process and the global search stochastic process. It is possible to define the local stochastic process as follows.

$$x_i^{(t+1)} = x_i^{(t)} + \alpha s \oplus H(p_a - \varepsilon) \otimes \left(x_j^t - x_k^t\right) \tag{8}$$

Where, ε is a random number drawn from the random distribution, $H(\cdot)$ is the Heaviside function, x_j^t and x_k^t are two separate random sequences, and s is the step size. The Lévy flight process is used to implement the global stochastic process.

$$x_i^{(t+1)} = x_i^{(t)} + \alpha \frac{\lambda\Gamma(\lambda) sin(\pi\lambda/2)}{\pi} \frac{1}{s^{1+\lambda}},\ s \gg s_0 > 0, 1 < \lambda \leq 3 \tag{9}$$

Where, the step size's scaling factor is defined as α.

Differential Evolution Algorithm (DE): The DE algorithm is the most potent random real-parameter optimisation method available today. It is a straightforward and effective global optimisation technique in continuous space. The parameter vector, mutation, initialization, selection and crossover are the first four processes in each iteration of the DE. An initial real parameter vector made up of NP individuals with d-dimension makes up the DE population. The following describes the expression of the population P^g and $\vec{x}_i^g$:

$$P^g = \left\{\vec{x}_1^g, \vec{x}_2^g, \dots, \vec{x}_n^g\right\} \tag{10}$$

$$\vec{x}_i^g = \left[x_{1,i}^g, x_{2,i}^g, \dots, x_{n,i}^g\right] \tag{11}$$

By using the differential mutation approach described below, the donor vector $\vec{v}_i^g = \left[v_{1,i}^g, v_{2,i}^g, \dots, v_{D,i}^g\right]$ corresponding to the i -th target vector $\vec{x}_i^g$ is produced:

$$\vec{v}_i^g = \vec{x}_{r1}^g + F \cdot \left(\vec{x}_{r2}^g - \vec{x}_{r3}^g\right) \tag{12}$$

Where, $F \in [0.4,1]$ defines the proportionate coefficient and, $\vec{x}_{r1}^g$, $\vec{x}_{r2}^g$, and $\vec{x}_{r3}^g$ are independent parameter vectors selected at random from the current population. Using a binomial crossover, the trial vector corresponding to the i-th target vector $\vec{x}_i^g$ is produced as follows:

$$u_{j,i}^g = \begin{cases} v_{j,i}^g, & rand_{i,j}[0,1] \leq Cr\ or\ j = j_{rand} \\ x_{j,i}^g, & otherwise \end{cases} \tag{13}$$

Where, Cr stands for the crossover rate, $rand_{i,j}[0,1]$ denotes the randomly chosen index value, and rand $j_{rand} \in [1,2,\dots,D]$ is a random integer subject to a uniform distribution. The next stage of the algorithm determines whether the target vector or the test vector should be kept for the following generation to maintain the number of the offspring population constant.

$$\vec{x}_i^{g+1} = \begin{cases} \vec{u}_i^g, & f\left(\vec{u}_i^g\right) \leq f\left(\vec{x}_i^g\right) \\ \vec{x}_i^g, & f\left(\vec{u}_i^g\right) > f\left(\vec{x}_i^g\right) \end{cases} \tag{14}$$

The variable $f(\cdot)$ stands for the minimised objective function in this situation. The algorithms DE and CS are excellent for optimisation. Although they are more rapidly convergent and have robust global search capabilities, a single optimisation model is effortlessly prone to local optimisation. As a result, the MCSDE is suggested as a novel exchange mechanism. In facilitating the transmission of information across both populations and limiting the development of local optima, this method increases the worldwide and local search capabilities of the optimization algorithm.

(ii) Stacked LSTM Model: A stacked LSTM architecture is a type of LSTM model that consists of many LSTM layers. Instead of sending a single value to the bottom LSTM layer, the higher LSTM layer sends a sequence. The model of accuracy is improved by the use of stacked LSTM, a stable method for solving difficult sequence prediction problems. A three-layered LSTM model is utilised in this work to fit the data. The LSTM model's training and prediction processes can be broken down into the following three parts. As the magnitude of the input data had an impact on the presentation of the LSTM models, the data were first rescaled and normalised to the range of 0 to 1. Instead of training several LSTM models for various values of n, the model is trained to predict the value for the following day since it predicts future n number of days pollutants. The time steps of the multivariate and univariate LSTM are set to 20/70, which implies that the incidence of the following day will be predicted using the data from the past 20/70 days. A three-layer layered LSTM structure was subsequently created. Every LSTM layer has an unseen layer with the neuron choices 6/40/90 that was pre-set for the LSTM model.

By accurately forecasting six daily air pollution concentrations (PM2.5, PM10, SO_2, NO_2, CO and O_3), the proposed unique hybrid forecasting technique evaluates total air quality, can and provide system-based air quality prediction early warning. Work must be done in terms of prevention and control to reduce air pollution and encourage the production of clean air. Developing air quality early warning systems will effectively monitor air pollution levels to reduce air pollutant emissions and encourage the development of cleaner production.

5 EXPERIMENTATION AND RESULT DISCUSSION

In this section, the proposed model's predictive power is evaluated using real data sets from a variety of Indian regions, including Kannur, Calicut, Palakkad, and Coimbatore in Southern India. For the current study locations, the risk analysis of PM10 and PM2.5 was evaluated using the AQI. The index for the daily and hourly average concentration was calculated using the revised WHO recommendation as the benchmark. By comparing data gathered before and after signal organization plans were implemented, the experiment's primary goal was to investigate the impact of signal coordination on vehicle emissions. The first stage in this pilot project was to gather sufficient data to characterise run variability, which would subsequently help determine the minimum number of runs required to produce statistically meaningful comparisons. The quantity of runs and vehicles employed heavily depends on the study's objectives. Studies with various goals would have different designs. For instance, more vehicles would be used but fewer runs would be made per vehicle in research to characterise fleet average emissions.

PM2.5, NO_2, SO, CO, and ozone were among the meteorological features that were collected hourly from 4 stations across India. The data was gathered during one year, from July 1, 2022, to June 30, 2023. The specifics of the implemented data sets and the datasets used for analysis can store data that can be utilised by a system program. The applications, such as system variables or parameters, macro libraries, or source programs, which is also required to store data sets. For both the training and testing sets of data, predictions are made using the stacked LSTM model to evaluate the model's performance.

Table 1 Pollutant Concentration of Ambient Air on RN

S. No	Measurement Date	Pollutant Concentrations					Unit
		CO	NO_2	SO_2	PM_{10}	$PM_{2.5}$	$\mu g/Nm^3$
1	15.06.2022	10041	110	135	188	99	$\mu g/Nm^3$
2	29.08.2022	10117	105	142	97	171	$\mu g/Nm^3$
3	17.10.2022	10345	96	91	116	136	$\mu g/Nm^3$
4	02.01.2023	10205	117	121	158	275	$\mu g/Nm^3$
5	24.04.2023	11009	128	114	241	163	$\mu g/Nm^3$
6	26.06.2023	10849	136	138	167	301	$\mu g/Nm^3$

Tab. 1 shows information on ambient air pollutant concentrations caused by traffic in the South Indian cities of Kannur, Calicut, Palakkad, and Coimbatore. It measures the AAQ y and displays the values of CO, NO_2, SO_2, PM10, and PM2.5 from October 2022 to December 2022, accordingly.

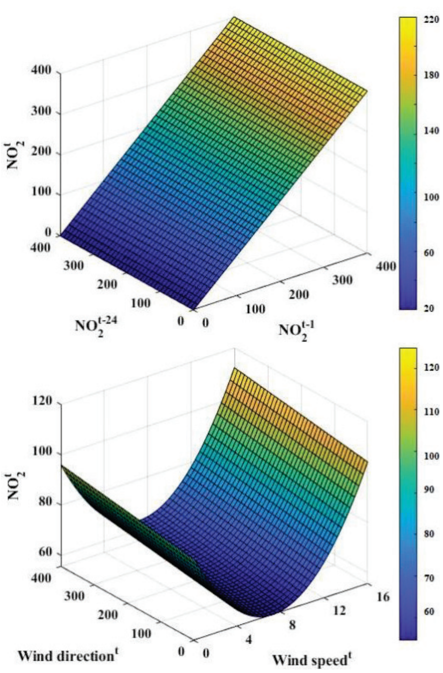


Figure 2 Predicted Results of NO_2 Timeline and Wind

The results for NO_2 in Ambient Air Quality measurements are depicted in Fig. 2. The wind direction input used an algorithm to calculate the half-hourly representative wind from wind readings in the monitoring stations. As a result, the selected city's use of coal and oil also tended to decrease, and the level of NO_2 in the air similarly decreased.

Based on the analysis of traffic volume, vehicle speed, and parking occurrences, Fig. 3 shows the traffic flow rate results for this task. The traffic flow estimation is primarily reliant on historical and current traffic information gathered from real-time sensors. The forecast is intended to be made for a certain day, and the traffic profile matches that day. However, these traffic profiles were chosen from a pool of archival information and shared trends with the subject profile. This period is taken into account when comparing the applicant and subject profiles.

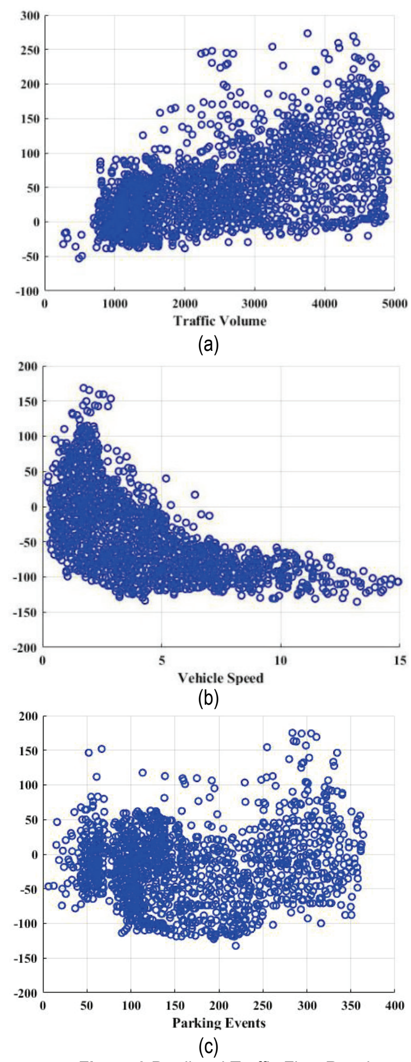


Figure 3 Predicted Traffic Flow Results

Fig. 4 illustrates the outcomes of the hybrid forecasting models' predictions of air pollution. Chemiluminescence technique is used in these monitoring stations to determine NOx content. Calibration problems, power outages and other technical problems prevented the stations from recording concentrations during all 17,520 half-hour time points.

However, at that time, data from more than 35 sites was accessible.

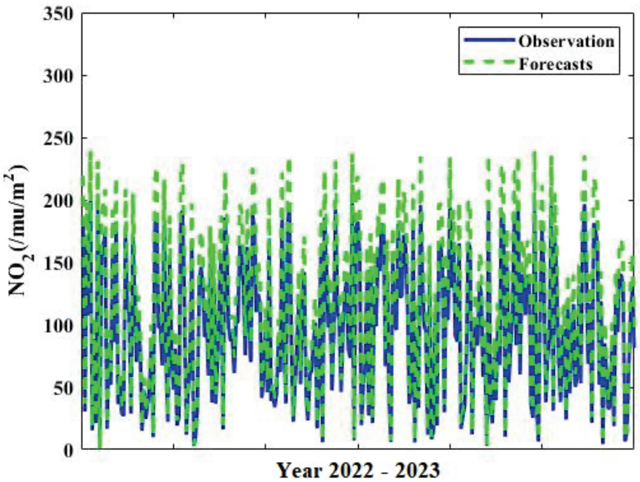


Figure 4 Ambient Air Pollutant Concentrations: Forecast and Observed Results

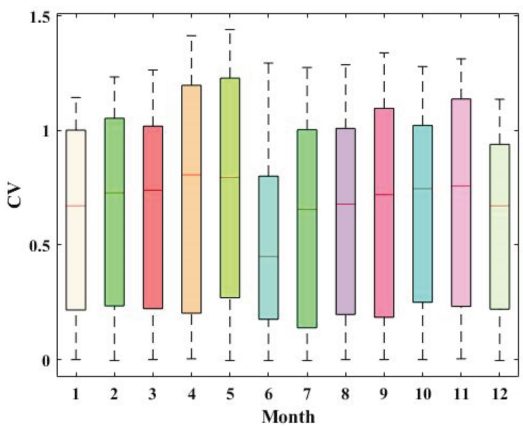
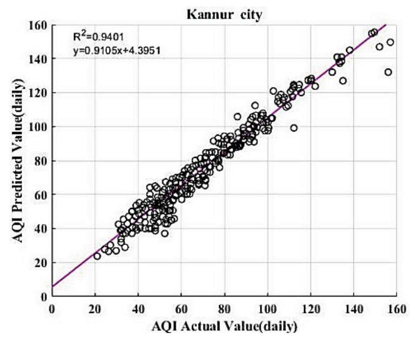


Figure 5 The coefficient of variation (CV) curve of NO_x concentration was modelled

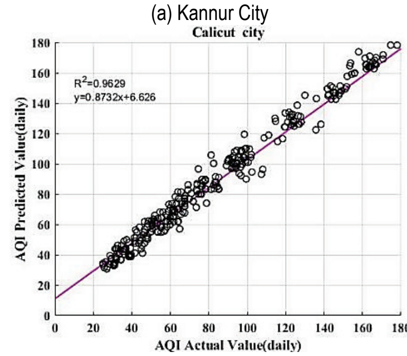
Fig. 5 displays the CV graphs of the modelled NO_x concentrations for the year 2022. When considering the temporal activity patterns of individuals or exposures during particular periods, the granularity displayed by this model is anticipated to have an impact on exposure assessment. Even though it lacks direct input data on atmospheric stability, it complies well with the local climatic patterns.

The regression analysis graph for AQI prediction in the cities of Kannur, Calicut, Palakkad, and Coimbatore is shown in Fig. 6. The trend break-line chart and regression analysis chart of these cities for the year from July 2022 to June 2023 are used in the prediction stage. This graph represents the majority of pollutant performance metrics and AQI values. For this AQI index prediction, the suggested hybrid forecasting model will produce superior results.

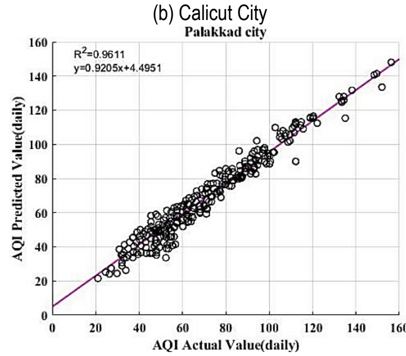
Fig. 7 shows the association between PM 10 and PM 2.5 for the 0.05 significant level for Kannur (r = 0.9959, P = 0.05), Calicut (r = 0.9960, P = 0.05), Palakkad (r = 0.9962), and Coimbatore city (r = 0.9916). The PM 10 vs. PM 2.5 scatterplot shown in Fig. 4 is also very tightly distributed about an underlying straight line. A clear linear connection can be seen in the quantified correlation between PM 2.5 and PM10 for Benin City. To simulate the more dangerous PM 2.5's likely source location and dispersion patterns in the analyzed sites, a model of hybrid forecasting was utilized. For the study, the peak PM 2.5 values from the months with the greatest level of particulate material (by season and region) were taken into account. The linear link between the particulate matters indicated by the straight line suggests that the level of PM2.5 rises as the level of PM10 rises. Furthermore, comparable bases of PM10 and PM2.5 may be used to explain both the association and the tightly distributed pattern.



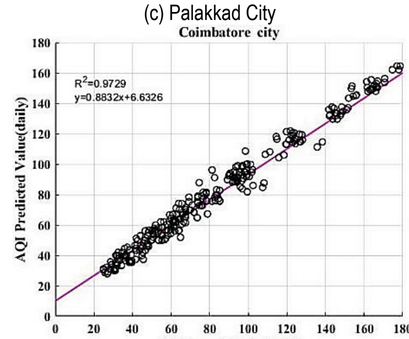
(a) Kannur City



(b) Calicut City

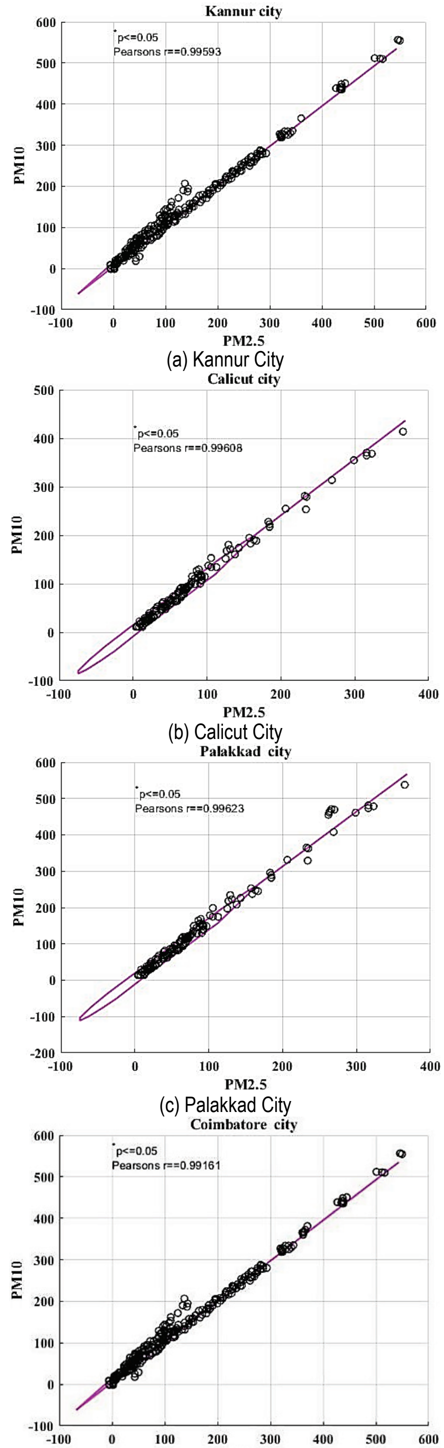


(c) Palakkad City



(d) Coimbatore City

Figure 6 Chart of Regression Analysis for Prediction of AQI



(a) Kannur City

(b) Calicut City

(c) Palakkad City

(d) Coimbatore City

Figure 7 Pearson correlation coefficient between PM2.5 and PM10 particles

The AQI plot based on daily collected data is shown in Fig. 8. The AQI was used to assess the risk analysis of PM10 and PM2.5 for the current research regions. The new WHO recommendation was used as the norm for calculating the index for the daily average concentration.

The performance findings for the concentrations of ambient air pollutants are displayed in Tab. 2. It displays the SO_2, CO, O_3, PM10, PM2.5, and NO_2 MAE, RMSE, and R^2 values. For CO and SO_2, it yields the lowest MAE, RMSE, and R^2 values of 0.048, 0.109, and 0.579, respectively.

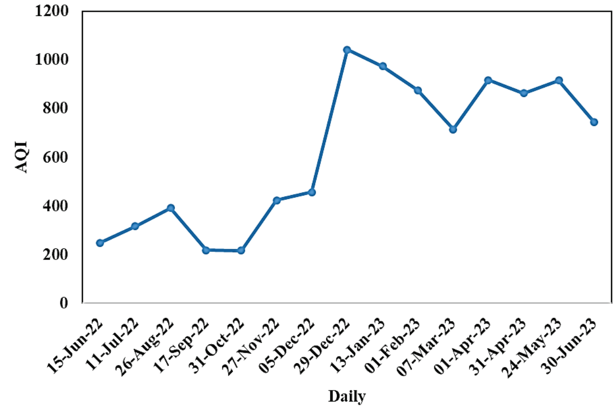


Figure 8 AQI Plot for Daily Measures

Table 2 Performance Results for AAP Levels

Pollutant	MAE	RMSE	R^2
SO_2	0.739	1.586	0.579
CO	0.048	0.109	0.793
O_3	3.846	5.549	0.905
PM_{10}	4.830	7.445	0.640
$PM_{2.5}$	3.359	4.670	0.843
NO_2	2.451	3.638	0.833

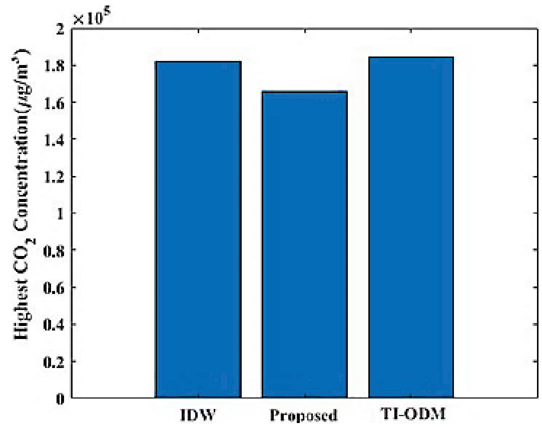


Figure 9 Comparison Results for the Highest Attention of Carbon Dioxide

Fig. 9 shows the comparison outcomes for the greatest carbon dioxide (CO_2) concentration. When compared to the current IDW and TI-ODM, the suggested technique yields less accurate findings than the others. As a result, the proposed approach reduces the attentiveness of air pollutants brought on by vehicle exhaust on the designated RN more than other existing approaches.

The comparison graph for the nitrogen oxide (NO_x) concentration with the highest value appears in Fig. 10. The outcome is contrasted with the current IDW and TI-ODM techniques. For nitrogen oxide (NO_x), the overall emission value has decreased from the current situation by 327.387 gr for the suggested approach and 562.601 gr for the existing work.

The cross-validated correlation coefficient between measurements of ambient NO_x and model values is shown in Fig. 11. The proposed method outperformed the IDW and TI-

ODM interpolation in terms of mean spatial Pearson correlation (MSPC) when compared to these existing techniques. This shows that, compared to the other investigated models, the suggested hybrid optimisation gives better accounts for the spatial patterns of NO_x. It is notably clear that the suggested work has better MSPC than TI-ODM. Given that midday strong vertical turbulence swiftly disperses traffic emissions, this may be yet another effect of the ODM's decoupling of the traffic and industry terms.

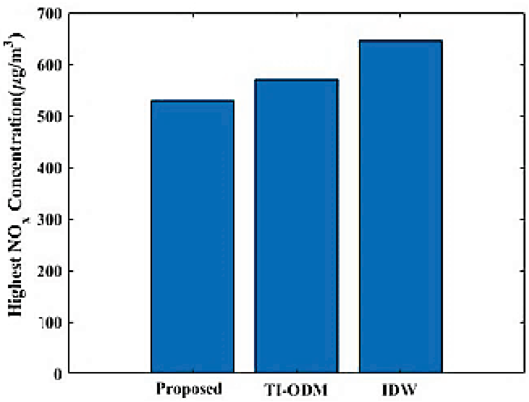


Figure 10 Results of Comparison for the Optimal Nitrogen Oxide Concentration

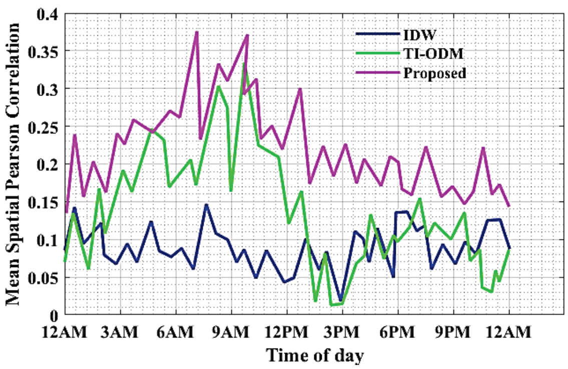


Figure 11 Measurements of Ambient NO_x and Cross-Validated Correlation Coefficient between Modelled Values

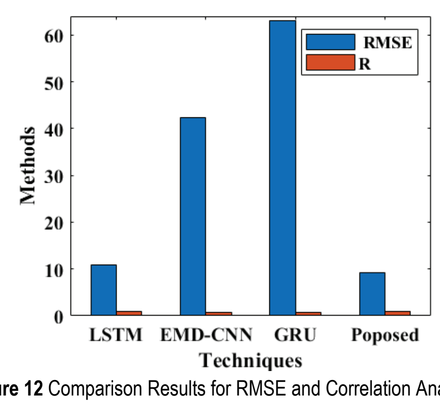


Figure 12 Comparison Results for RMSE and Correlation Analysis

The comparison between the findings of RMSE and regression analysis is shown in Fig. 12. It demonstrates that the RMSE of various approaches lowers as the depth increases, but at a particular depth, an inflexion point occurs, and the RMSE abruptly increases. The suggested approach is contrasted with the current LSTM, EMD-CNN, and GRU models in this graph. The proposed technique yields higher R values and the RMSE value is very low in this figure, respectively.

The RMSE and Pearson coefficient compared to forecasted findings are displayed in Tab. 3. Here, the suggested approach yields higher R values of 0.98 and RMSE lower values of 9.17, respectively. Therefore, the proposed method outperforms the other available methods in terms of performance.

Table 3 Comparative Predictive Analysis

Techniques	*RMSE*	*R*
LSTM	10.9	0.97
EMD-CNN	46.26	0.81
GRU	63.1	0.65
Proposed	9.17	0.98

6 RESEARCH CONCLUSION AND FUTURE SCOPE

Air quality is known to have a considerable impact on health, predicting it is a very essential undertaking. More accurate and complex modelling techniques are required as environmental regulations become stricter to simulate measures and programmes that may successfully address air quality exceedances, which are frequent in major Indian cities, particularly for NO_2. Since observed concentration values are the result of the interaction of several sources and processes spanning a broad range of geographical and temporal dimensions, modelling air quality in metropolitan settings is a relatively challenging task. Measurements of outdoor concentrations and individual exposure to CO, NO_2, SO_2, and O_3 will be made as part of this study using the real-time sensor and a pilot campaign. Traffic data in the form of traffic volume, vehicle speed, and number of parking times were collected by a pilot survey at the study site. Air pollution data was measured on-site with support from the Road Transport Research Technical Enforcement Unit. For each road segment, traffic flow is recorded every hour. AAP data were collected using an impact tube to measure NO_2, SO_2, and O_3 pollutants, a non-dispersive infrared instrument to measure CO pollutant concentrations, and an HVAS instrument to measure PM10 pollutant concentrations. The most common method for estimating past exposures to air pollution is regression modelling, such as Land-Use Regression. To forecast AAQ levels, a GDM with a residual network model is provided.

However, using modules for air quality evaluation and prediction, the study develops a trustworthy early-warning system. A hybrid forecasting technique is introduced in the forecast module to anticipate pollution concentrations that accurately predict future air quality conditions. A pollution event happens; this model is used to develop an alarm protocol (one day in advance). The results were compared with ambient air measurements at the study site and alternative traffic-related solutions were proposed. The results show that although the forecast model outperforms other standard models in predicting pollutant concentrations, the assessment model is only suitable for reporting air quality levels compared to the actual situation. Therefore, the proposed system is expected to play an important role in developing smart cities and reducing air pollution globally in the future. According to the test results, the newly developed SAP is significantly more suitable for air pollution research and monitoring, thus adding a new and viable alternative for decision makers. Furthermore, a health risk assessment was proposed to determine the risk sensitivity of PM10 and PM2.5 in the five districts of NEOM City. The results show that, compared with current state-of-the-art models, the proposed method with effective feature extraction can significantly optimize the accuracy of spatial air quality forecast-time. The MASE PM10 and PM2.5 values for the hour's prediction tasks were 97 and 99, respectively. Consequently, the suggested work is compared to another previous approach; in contrast to these existing methods, the proposed method yields greater correlation values, higher anticipated results, and a lower error rate. Consequently, with the early prediction and warning systems, the suggested approach offers a practical way to increase the level of alert for air pollution concentrations on India's RN while also being transferable to other regions around the world.

Future research in the field of forecasting daily Air Quality Index (AQI) and implementing Early Warning Systems (EWS) for ambient air pollution on road networks using the proposed Gaussian Dispersion Model with Deep Learning Algorithm could explore several promising directions. Firstly, investigating the integration of additional environmental factors, such as topography and land use, into the model may enhance its accuracy and broaden its applicability to diverse urban environments.

Furthermore, the exploration of ensemble methods that combine multiple forecasting models, including alternative dispersion models and machine learning algorithms, could offer a more robust and comprehensive approach to air quality prediction. Additionally, research focusing on real-time sensor data integration and the development of adaptive models capable of continuously learning from incoming data streams could further improve the system's responsiveness to dynamic and evolving pollution scenarios. Finally, assessing the scalability of the proposed methodology to larger urban areas and its potential integration with smart city initiatives would contribute to its practical implementation in real-world urban planning and pollution management strategies.

7 REFERENCES

[1] Wang, Y., Jin, X., Liu, Z., Wang, G., Tang, G., Lu, K. & Zhang, Y. (2023). Progress in quantitative research on the relationship between atmospheric oxidation and air quality. *Journal of Environmental Sciences*, *123*, 350-366. https://doi.org/10.1016/j.jes.2022.06.029

[2] Shen, L., Lu, X., Huynh, T. L. D. & Liang, C. (2023). Air quality index and the Chinese stock market volatility: Evidence from both market and sector indices. *International Review of Economics & Finance*, *84*, 224-239. https://doi.org/10.1016/j.iref.2022.11.027

[3] Liang, Q., Liu, S., Yin, J., Han, Q., Zhang, W. & Niu, J. (2023). Spatial–Temporal Characteristics and Influencing Mechanisms of Air Quality Index by Considering COVID-19 in Yunnan, Southeastern Tibetan Plateau. *Atmosphere*, *14*(2), 378. https://doi.org/10.3390/atmos14020378

[4] Bañeres, D., Rodríguez-González, M. E., Guerrero-Roldán, A. E., & Cortadas, P. (2023). An early warning system to identify and intervene online dropout learners. *International Journal of Educational Technology in Higher Education*, *20*(1), 1-25. https://doi.org/10.1186/s41239-022-00371-5

[5] Kołacz, A. M., Wiśnik-Sawka, M., Maziejuk, M., Natora, M., Harmata, W., Rytel, P. & Gajda, D. (2023). Air Pollution and Radiation Monitoring in Collective Protection Facilities. *Sensors*, *23*(2), 706. https://doi.org/10.3390/s23020706

[6] Ma, P., Zhou, N., Wang, X., Zhang, Y., Tang, X., Yang, Y. & Wang, S. (2023). Stronger susceptibilities to air pollutants of influenza A than B were identified in subtropical Shenzhen, China. *Environmental Research*, *219*, 115100. https://doi.org/10.1016/j.envres.2022.115100

[7] Seydou, T. H., Agali, A., Aissatou, S., Seydou, T. B., Issaka, L. & Ibrahim, B. M. (2023). Evaluation of the Impact of Seasonal Agroclimatic Information Used for Early Warning and Farmer Communities' Vulnerability Reduction in Southwestern Niger. *Climate*, *11*(2), 31. https://doi.org/10.3390/cli11020031

[8] Ghosh, S. & Mandal, R. (2023, January). Context Aware City Air Quality Monitoring Estimation. In *Proceedings of the 24th International Conference on Distributed Computing and Networking* (pp. 390-395). https://doi.org/10.1145/3571306.3571441

[9] Elbaz, K., Hoteit, I., Shaban, W. M. & Shen, S. L. (2023). Spatiotemporal air quality forecasting and health risk assessment over smart city of NEOM. *Chemosphere*, *313*, 137636. https://doi.org/10.1016/j.chemosphere.2022.137636

[10] Babaan, J., Hsu, F. T., Wong, P. Y., Chen, P. C., Guo, Y. L., Lung, S. C. C., ... & Wu, C. D. (2023). A Geo-AI-based ensemble mixed spatial prediction model with fine spatial-temporal resolution for estimating daytime/nighttime/daily average ozone concentrations variations in Taiwan. *Journal of Hazardous Materials*, *446*, 130749. https://doi.org/10.1016/j.jhazmat.2023.130749

[11] Ayodele, E., Okolie, C., Akinnusi, S., Mbu-Ogar, E., Alani, R., Daramola, O. & Tella, A. (2023). An assessment of the spatio-temporal dynamics of Landsat-derived aerosol concentration in relation with land cover and road networks in the Lagos megacity. *Environmental Science and Pollution Research*, *30*(15), 43279-43299. https://doi.org/10.1007/s11356-022-25042-w

[12] Bikis, A. (2023). Urban Air Pollution and Greenness in Relation to Public Health. *Journal of Environmental and Public Health*, *2023*. https://doi.org/10.1155/2023/8516622

[13] Ramentol, E., Grimm, S., Stinzendörfer, M. & Wagner, A. (2023). Short-Term Air Pollution Forecasting Using Embeddings in Neural Networks. *Atmosphere*, *14*(2), 298. https://doi.org/10.3390/atmos14020298

[14] Shanmugasundaram, P. & Bhatnagar, S. (2022, February). Co-operative Multi-Agent Twin Delayed DDPG for Robust Phase Duration Optimization of Large Road Networks. In *International Conference on Agents and Artificial Intelligence* (pp. 122-142). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-22953-4_6

[15] Zhang, Q. S. (2023). Environment pollution analysis on smart cities using wireless sensor networks. *Strategic Planning for Energy and the Environment*, 239-262. https://doi.org/10.13052/spee1048-5236.42112

[16] Sohrab, S., Csikós, N. & Szilassi, P. (2022). Connection between the spatial characteristics of the road and railway networks and the air pollution (PM10) in urban–rural fringe zones. *Sustainability*, *14*(16), 10103. https://doi.org/10.3390/su141610103

[17] Song, J. & Stettler, M. E. (2022). A novel multi-pollutant space-time learning network for air pollution inference. *Science of the Total Environment*, *811*, 152254.

https://doi.org/10.1016/j.scitotenv.2021.152254
[18] Galán-Madruga, D. & García-Cambero, J. P. (2022). An optimized approach for estimating benzene in ambient air within an air quality monitoring network. *Journal of Environmental Sciences*, *111*, 164-174. https://doi.org/10.1016/j.jes.2021.03.005
[19] Asha, P., Natrayan, L. B. T. J. R. R. G. S., Geetha, B. T., Beulah, J. R., Sumathy, R., Varalakshmi, G. & Neelakandan, S. (2022). IoT enabled environmental toxicology for air pollution monitoring using AI techniques. *Environmental research*, *205*, 112574. https://doi.org/10.1016/j.envres.2021.112574
[20] Martín-Baos, J. Á., Rodriguez-Benitez, L., García-Ródenas, R. & Liu, J. (2022). IoT based monitoring of air quality and traffic using regression analysis. *Applied Soft Computing*, *115*, 108282. https://doi.org/10.1016/j.asoc.2021.108282
[21] Kuenen, J., Dellaert, S., Visschedijk, A., Jalkanen, J. P., Super, I. & Denier van der Gon, H. (2022). CAMS-REG-v4: a state-of-the-art high-resolution European emission inventory for air quality modelling. *Earth System Science Data*, *14*(2), 491-515. https://doi.org/10.5194/essd-14-491-2022
[22] Wu, C. L., Song, R. F., Zhu, X. H., Peng, Z. R., Fu, Q. Y. & Pan, J. (2023). A hybrid deep learning model for regional O3 and NO2 concentrations prediction based on spatiotemporal dependencies in air quality monitoring network. *Environmental Pollution*, *320*, 121075. https://doi.org/10.1016/j.envpol.2023.121075
[23] Moses, O. & Ehinomen, A. B. (2023). Facile Ambient Air Quality Indicator Monitoring Technique for Estimating Cadmium and Lead Content in Atmospheric Dust. *Ethiopian Journal of Science and Sustainable Development*, *10*(1), 20-27. https://doi.org/10.20372/ejssdastu:v10.i1.2023.540
[24] Shin, H. & Bithell, M. (2023). TRAPSim: An agent-based model to estimate personal exposure to non-exhaust road emissions in central Seoul. *Computers, Environment and Urban Systems*, *99*, 101894. https://doi.org/10.1016/j.compenvurbsys.2022.101894
[25] Zhong, J., Hood, C., Johnson, K., Stocker, J., Handley, J., Wolstencroft, M. & Bloss, W. J. (2021, October). Modelling Street-Scale Resolution Air Quality for the West Midlands (UK) Using the ADMS-Urban RML System. In *International Technical Meeting on Air Pollution Modelling and its Application* (pp. 77-82). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-12786-1_10
[26] Neelaveni, N. & Rajeswari, S. (2016). Data mining in agriculture-a survey. *International Journal of Modern Computer Science—Revista da Faculdade de Serviço Social da UERJ, Rio de Janeiro*, *4*(4), 104-107.

Authors' contacts:

Asha Unnikrishnan, Research Scholar
(Corresponding author)
Department of Computer Science, Sree Saraswathi Thyagaraja College,
Palani Road, Pollachi, Tamil Nadu 642107, India
E-mail: asha.nkk@gmail.com

S. Rajeswari, Associate Professor, Dr.
Department of Computer Science, Sree Saraswathi Thyagaraja College,
Palani Road, Pollachi, Tamil Nadu 642107, India
E-mail: rajeswari@stc.ac.in