

Predicting the least air polluted path using the neural network approach

K.Krishna Rani Samal^{1,*}, Korra Sathya Babu¹, Santos Kumar Das¹

¹National Institute of Technology, Rourkela, India

Abstract

Air pollution exposure during daily transportation is becoming a critical issue worldwide due to its adverse effect on human health. Predicting the least air polluted healthier path is the best alternative way to mitigate personal air pollution exposure risk. Computing the least polluted path for the current time might not be helpful for real-time applications. Therefore, we develop a routing algorithm based on a neural network-based CNN-LSTM-EBK (CLE), a temporal-spatial interpolation model. The proposed model predicts pollution levels at high temporal granularity. This paper introduces a weight function to compute air pollution concentration at the road network. It also predicts the least air polluted path among all possible paths from a source to a destination at different time granularity. The results show that the predicted path may be longer than the shortest route but minimize pollution exposure risk all the time, which proves its effectiveness during daily transportation.

Received on 06 April 2021; accepted on 19 June 2021; published on 29 June 2021

Keywords: Air quality modelling, Routing, Deep learning, GIS, Kriging

Copyright © 2021 K.Krishna Rani Samal *et al.*, licensed to EAI. This is an open access article distributed under the terms of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>), which permits unlimited use, distribution and reproduction in any medium so long as the original work is properly cited.

doi:10.4108/eai.29-6-2021.170250

1. Introduction

The unpredictable development of urbanization in recent years has resulted in many undesirable environmental challenges. The unavoidable rising of ambient air quality level is one of them [1, 2]. In past years, it is observed that a higher concentration of air pollution arises near the road network, which affects pedestrians, drivers, school children, older people, and patients during their travel. Traffic-related or vehicular-related air pollution is one of the main reasons behind this increasing level of outdoor air pollution and the deterioration of public health conditions. PM_{10} (particulate matter with diameter 10 micrometers) is considered as the major harmful pollutant among all the pollutants due to its severe negative impact on the environment and human health [3]. It can affect drivers and pedestrians, causing many heart diseases, asthma, and respiratory problems [4]. People perform physical activities such as walking to improve their health condition, but sometimes it becomes worse due to the road network's unhealthy air pollution level.

In order to get rid of this problem, the ultimate solution is to predict the air pollution level in the road network and predict the least air pollution exposure path amongst all possible paths to provide an alternative safe route navigation system. Predicting the least air polluted path during walking, cycling, and driving can minimize pollution exposure risk during daily transportation. The first step in predicting the least air-contaminated path is the continuous pollutant concentration mapping. In order to get a proper surface mapping of air quality, meteorological and road network datasets can be collected and model a better routing solution. This method can be developed by performing temporal, spatial analysis of data. Subsequently, based on those temporal, spatial prediction models, a pollutant's continuous surface can be achieved. Air pollutant value prediction based on past observations data in the real-world scenario is challenging. Pollution level prediction at road networks and rank those paths according to their pollution level is one of the most difficult tasks for citizens due to the unavailability of accurate air pollution data at each geographical point. The limited and sparsely distributed air pollution monitoring stations are the

*Corresponding author. Email: 517cs6019@nitrrkl.ac.in

main reason behind this issue. As air quality has a temporal and spatial dependency with the increasing air pollution level, it is challenging to analyze both the dependence and predict the air pollution level at the road network's unmeasured locations.

Predicting air pollution level includes several types of methods e.g., statistical [5–8], deterministic, numerical [9], geostatistics, machine learning [10], and deep learning techniques [11]. The existing techniques are not efficient enough to predict air pollution levels at the road network at different time granularity. Therefore, this paper proposes a deep learning-based interpolation methodology framework to mitigate the limitation of the existing prediction models and compute the road network's air pollution level at different temporal resolutions. This paper utilized the PM_{10} pollutant concentration and open street map of Odisha to evaluate the proposed methodology framework's performance. Besides, the interpolated PM_{10} value is treated as a cost function to predict the shortest path from starting to an endpoint.

Following the introduction, the rest of this paper is organized as follows: Section 2 formulates the problem which motivates this research work. Section 3 gives an idea about the literature survey. Section 4 and 5 explain the data source and methodology used for the proposed routing algorithm. Section 6 shows the experimental results, and Section 7 briefly describes the conclusion drawn from the experimental results.

2. Problem Formulation

Let $G \hat{=} (V, E)$ is the undirected graph, where the graph G is having N number of nodes V . Vertices $(v_i, v_j) \in V$ and edge $(e_{ij}) \in E$ between two vertices is often used for spatially modelling the geographical data. Each air pollution monitoring stations V on G , has time series pollution data for time t with features f . These input features can be implemented in G as a feature matrix $F_t \in R^{N \times f}$. The road network feature f can be any pollutant value. The problem of spatial temporal analysis for road network is to develop a weight assignment mapping function w on road network G to predict f value in the next d days as mentioned below: $(F_{t+1}, F_{t+2}, F_{t+3}, \dots, F_{t+d}) = w(G, (F_{t-n}, \dots, F_{t-2}, F_{t-1}, F_t))$. Where n is the length of the historical time series dataset. The spatial relation among air pollution monitoring stations can be represented as an undirected graph $G \hat{=} (V, E, w)$. The next problem is to find the best alternative path R from source s to any destination h node on $G \hat{=} (V, E, w)$ which can be considered as the healthier path.

3. Literature survey

Many smart initiatives are making citizens' life much more comfortable and safer. One of the smart initiatives of the government is to improve the health conditions and wellness of people. Deterioration of air pollutants has become major trouble towards these initiatives. The government is taking necessary actions like personalized air pollution management, air quality monitoring, air quality modeling, and developing a smart transportation system with the minimum air pollution exposure risk to manage public health. Many IoT devices are installed at many locations, and pollution monitoring stations are established to perform all these activities. These devices have been installed to collect air pollution levels across many geo-locations to identify the pollution level at that particular location [4]. Still, those sensors are unable to predict air pollution levels on road networks. Existing personalized routing solution is based on user's preference like traffic condition [12], shortest [13], fastest [13–15], the least expensive path, lower crime [16–18] and lower accident-prone path [19] [20] but not on the air pollution exposure level. Nowadays, one has to be more concerned about adverse air pollution [21] and weather [22] condition impact on the road network for safer travel. Personalized routing solution, part of personalized pollution management, is a critical aspect of air pollution management. Exposure to air pollution for a longer time duration may cause severe health issues to older people, school children, and pregnant women [23]. Still, very few research has been conducted to develop a smart navigation system based on air pollution exposure risk [24–27].

This section reviews some state-of-the-art methods on safe route navigation that minimize air pollution exposure risk. Muller *et al.* conducted experiments using PM_{10} emission data of Berlin [24], as it is a major indicator of pollution. The proposed method suggested air quality-adjusted shortest path for cyclists and pedestrians. Sharker *et al.* [26] proposed a framework, which provides a health optimal routing solution by computing optimal health weight for various scenarios. The author implemented the Bayesian modeling and influence diagram technique to design a weighting model considering individual and environmental variables and recommended a health-optimal navigation service for the pedestrian. An IoT-based novel architecture is proposed in [27] to get the real-time routing solution. Air quality on the road network is of great importance to take precautionary measures for a healthier life, hence to make citizens aware of this, Ramos *et al.* [25] promotes pollution free routes based on the existing air quality sensor network. The technology-agnostic methodology framework is based on the air quality index interpolation and helps

trace the city's air pollution-free routes. A study was conducted in Taiwan [28] to compute air pollution exposure levels during different transportation modes like walking, motorcycle, and bicycle transportation mode. Zahmatkesh *et al.* [29] proposed a geostatistical interpolation technique, i.e., kriging to generate air quality index spatial prediction map to prevent the air pollution zone and its polluted routes.

All the existing work attempted to develop an intelligent navigation system is based on the spatial distribution [30] of past air pollution data, which practically does not significantly improve the quality of human health. Moreover, the existing work does not predict the air pollution level at the road network for the near future.

Therefore, it is better to develop an early warning navigation system, alerting people by predicting air pollution levels in the road network in advance. Based on the above survey, this paper developed an algorithm based on the temporal-spatial interpolation model to minimize personal air pollution exposure risk to improve public health conditions.

4. Data collection

The ambient air quality data of Odisha, India [31, 32] is used to evaluate the proposed model. The dataset includes day-wise sampled PM_{10} , $PM_{2.5}$, SO_2 , NO_2 pollutant information in micrograms (one-millionth of a gram) per cubic from 2005 to 2015. The dataset is having 11 years of data for 30 monitoring sites of Odisha. Due to many missing values, 14 sites' information is dropped, and 16 monitoring sites' data sets are utilized for evaluation purposes. It also contains station code, sampling date, monitoring station name, type of location, longitude, and latitude value as attributes. The study area's road network graph and boundary are collected from the Open street map to evaluate the proposed algorithms. From the descriptive analysis of past observations, it is found that PM_{10} is the major contributing air pollutant in Odisha [33–35], so utilized for experimental purposes to determine air pollution exposure risk in the study area.

5. Predicting the least PM_{10} exposure risk path

In this section, the step-by-step proposed methodology is explained to predict the path with a minimum of PM_{10} exposure. First, we integrate PM_{10} concentration with its spatial feature and road network of Odisha. The proposed methodology combines both time and space dimensions simultaneously to perform both spatial and temporal analysis and generate a prediction map of the study area at each time instance. The temporal-spatial interpolation is required to assign an average PM_{10} concentration value on the road network, which will

be treated as a weight function to compute the shortest path from a source to a destination.

5.1. Temporal-Spatial air quality modeling

A person may take one or two days to travel from a source to a destination. Hence, we first require identifying how PM_{10} concentration changes over time throughout the journey. To complete this task, we have to perform PM_{10} prediction in the road network for the next day or next week. The proposed architecture, which is developed to achieve this objective, is presented in Figure 1.

Experiments and discussions of each part of the proposed architecture are explained in the following subsection.

Data normalization. Before completing the data analysis steps, it is required to normalize the data and remove the dataset's existing outliers. The existing outliers may affect the model's performance; hence Z-Score normalization is conducted in the preprocessing step, followed by the removal of the outlier. The equation for Z-Score normalization [36] is given below,

$$Z - Score_i = \frac{\sum_{i=1}^n p_i - \hat{x}}{S} \quad (1)$$

where p_i is the pollutant value, \hat{x} represents mean and S represents the standard deviation. After performing the preprocessing step, outputs are utilized to perform temporal air quality modeling to get prediction value for the next few days.

Feature extraction. A convolutional neural network (CNN) [37] is used to extract the features from preprocessed normalized historical pollution dataset and identify the relationship between the features that might influence the increasing air pollution level. Weight sharing and sparse connectivity features enable CNN to handle multiple layers and nonlinearity. It updates the weights, which contributes to the prediction results. Thus, it reduces the number of parameters and improves performance. Then, the max-pooling operation of stride two is conducted in input to decrease the convolution size and hence, increases the computational complexity. The schematic diagram of the convolutional neural network is represented in Figure 2.

Temporal modelling. Extracted features from the CNN layer fed into the long short-term memory (LSTM) [11] network to capture the long-term and short-term dependency of air quality for sequential modeling. A selective write, read and forget strategy followed in the LSTM model to deal with longer sequences to perform air pollution prediction at low and high time

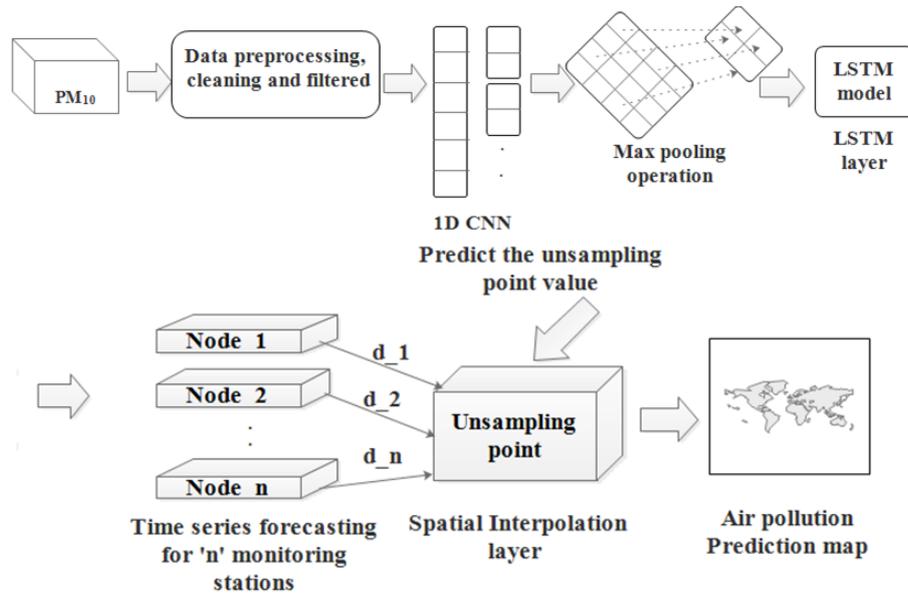


Figure 1. The proposed LSTM based temporal-spatial interpolation model architecture.

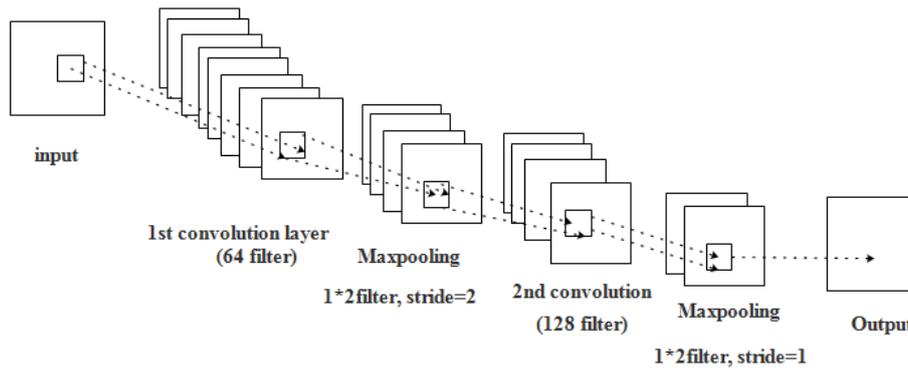


Figure 2. Schematic diagram of Convolutional neural network.

granularity. The basic schematic diagram of the LSTM model is represented in Figure 3.

The goal of this LSTM basic unit is to compute s_t . It can be mathematically formulated as,

$$s_t = f(s_{t-1}, x_t) \quad (2)$$

while computing s_t , it will take past information s_{t-1} and current input information x_t , in order to keep only relevant information. Instead of writing all parts of s_{t-1} , it can use a few parts of it for the next step, called selective write. o_t is used to decide how much of the information should be retained and pass to the next step. As the output from selective write, it may not be interested in reading all this information, so it will be calculated with an input gate i_t to perform a selective read operation. Then f_t is used to forget the unnecessary input information. After computing the selective write, read, and forget gate, it will compute some states, and all these operations will be guided by the write, read,

and forget operation. The computation of gates at time t can be represented mathematically as,

$$o_t = \sigma(w_o h_{t-1} + u_o x_t + b_o) \quad (3)$$

$$i_t = \sigma(w_i h_{t-1} + u_i x_t + b_i) \quad (4)$$

$$f_t = \sigma(w_f h_{t-1} + u_f x_t + b_f) \quad (5)$$

where, (o_t, i_t, f_t) are the output gate, input gate and forget gate respectively. σ is the sigmoid activation function, (w_o, w_i, w_f) are the hidden state parameters, (u_o, u_i, u_f) are the current state parameters and (b_o, b_i, b_f) are the bias parameters. The computation of states can be formulated as follows,

$$\hat{s}_t = \sigma(w h_{t-1} + u x_t + b) \quad (6)$$

$$s_t = f_t \odot s_{t-1} + i_t \odot \hat{s}_t \quad (7)$$

$$h_t = o_t \odot \sigma(s_t) \quad (8)$$

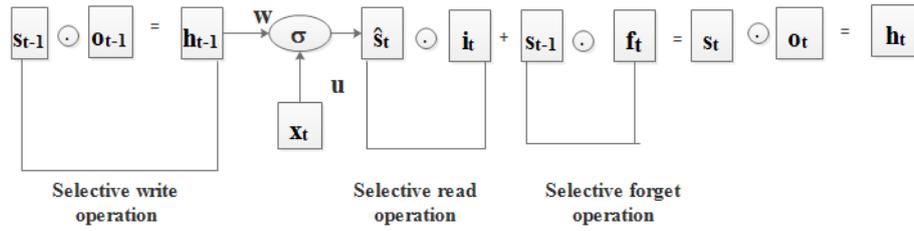


Figure 3. The basic schematic diagram of the LSTM. Selective write, read and forget operations are performed for all the time stamps t to retain some current and past information and get the final output h_t

where \hat{s}_t is the temporary state, s_t is the combination of current temporary value that computed in selective read and some forgotten part from the selective forget operation, h_t signifies how much of the information need to write in the next stage, which will be controlled by output gate o_t and the temporary state s_t . After performing all the operations, the backpropagation algorithm will have 12 parameters that will be learned. Now one can compute the derivative of the loss function concerning each of those parameters to reduce the loss function.

Temporal-spatial interpolation prediction. Each road network has different features like different road lengths, widths, and different road segments. Due to unstructured monitoring stations, it is challenging to compute the PM_{10} level at each road segment. In the past years, many works have been done to identify the spatial distribution of air pollution and estimate unknown point values in a particular study area using spatial interpolation techniques. Although these spatial interpolation techniques are widely used, many critical problems remain unsolved as the traditional spatial interpolation technique considers time and space individually. It may not be a satisfying method while predicting the spatial distribution in many practical applications at a different time granularity. Spatial interpolation for the next few days may add an advantage to make an efficient method that can alert people in advance to keep them safe. The most of the existing interpolation methods e.g. radial basis function (RBF) [38, 39], ordinary kriging (OK) [40–43] and inverse distance weighting (IDW) [44, 45] only predict the spatial distribution of air pollution for the current time not for the future duration. Thus, it is required to perform spatial interpolation based on temporal interpolation to predict interpolation results at each unsampled point and time instance.

All the spatial modeling techniques are based on the first law of geography, i.e., nearer points have similar characteristics having a stronger correlation, which includes IDW, kriging, and RBF interpolation techniques. IDW interpolation is a function of distance, which estimates the unsampled point values using a linear combination of sampling point values weighted by the inverse function of the distance from the

interpolation point to the sampling points. While kriging is the weighted average of the measured data, which is based on the semivariogram. But these kriging methods are not so accurate due to poor standard error value, which also assumes the intrinsic stationarity, but unfortunately, air pollution data are not stationary. RBF, a deterministic method similar to kriging techniques but does not benefit from spatial analysis using a variogram. This technique is used to generate continuous surfaces using a large number of data points [46, 47]. Unlike other interpolation techniques, the EBK (empirical Bayesian kriging) method does not require the prior distribution details. It can handle local and global data stationarity by providing faster performance with the default parameter setting [48]. This interpolation method is a different technique than any other existing traditional interpolation technique as it accounts for the error during the use of many semivariogram models instead of one. The conventional model uses one semivariogram model to adjust the parameters manually. The EBK method follows the subsetting and simulation process to adjust the parameters ϕ_k and develop a better model automatically. It follows three main steps while getting interpolation results. First, it estimates a semivariogram model and parameters ϕ_k using the observed data. Secondly, using this semivariogram and ϕ_k , new values can be estimated by doing simulations at each input location M times. Thirdly, newly simulated inputs can be used for generating a new semivariogram model. Bayes' rule can be used to assign a weight w_k to this newly estimated semivariogram model and new parameters ϕ_k . It can be formulated using the following formulas:

$$w_k \propto f(P|\phi_k) \quad (9)$$

$$w_k = \frac{f(P|\phi_k)}{\sum_{k=1}^M f(P|\phi_k)} \quad (10)$$

$$\sum_{k=1}^M w_k = 1 \quad (11)$$

where weight is presented as w_k , ϕ_k is the simulated parameters, P is the observed pollutant value and $f(P|\phi_k)$ is the conditional probability of the observed pollutant value P given the model parameter ϕ_k . It signifies how likely the observed PM_{10} value can be generated from the estimated model. The second and third steps are repeated to estimate the semivariogram model in step one, which can be used to simulate a new set of values at each input point. These simulated data can be used to estimate the new semivariogram model and its weights. These weights (w_k) are used to generate predictions and prediction errors at each unsampled point. So, this interpolation can be an effective spatial interpolation technique for the road network. Hence, this research work experimented with the EBK model's effectiveness by adding it to the top of the CNN-LSTM model to get the prediction result for $t+d$ days at the road networks. The prediction results of multiple nodes are used as input to the spatial interpolation layer to predict the spatial distribution of PM_{10} , where the nodes' coordinates are treated as the boundary of the study implementation area. Thus, adding time series prediction results as input to this EBK interpolation layer yields better prediction results. Finally, the output of the proposed model generates a temporal-spatial interpolation prediction map of the study area. The Pseudocode of the proposed algorithm described above is represented in the Algorithm 1, where the first part of the algorithm performs temporal modeling and the second part perform the spatial modeling.

The temporal modeling is the most crucial step of PM_{10} prediction because a journey may take more than one day. Sometimes, it requires predicting the PM_{10} value for the next week or the next month to make the right move. The second part of the algorithm is responsible for spatial interpolation. As the air pollution monitoring stations are in unstructured locations like some might close to each other or far away. Therefore, EBK interpolation is used to assign the average PM_{10} value as the weight on the road segment to predict each road segment's pollution level for the $t+d$ days, depending upon the arrival time on the destination.

5.2. Application of the proposed model for safe navigation

The proposed CNN-LSTM-EBK (CLE) model predicts the PM_{10} level at each location over the $t+d$ days in the study area effectively. So, the proposed model is utilized for the application of the transportation system in a smart city to predict the healthiest safest path to the citizens. The temporal-spatial interpolation prediction map generated by the CLE model is used as input to propose a routing algorithm, which reduces PM_{10} pollutant exposure risk. It may suggest a comparatively

Algorithm 1 LSTM based temporal-spatial interpolation prediction method for road network

Input: Air pollution historical data $P = [p_1, p_2, p_3, p_4, p_5, \dots, p_T]$, road network graph $G \hat{=} (V, E)$, latitude i , longitude j .

Output: Pollutant prediction surface $G \hat{=} (V, E, w)$

Initialization: Manually trained model (K) with parameters θ

- 1: **for** $1 \leq i, j \leq N$ **do**
- 2: **for** $t = 1$ to T **do**
- 3: **if** P has null value for time t **then**
- 4: do linear interpolation
- 5: **else**
- 6: $\hat{p}_{t+d}^{i,j} \leftarrow (i, j)$ Generate time series prediction result for V by K .
- 7: return $\hat{p}_{t+d}^{i,j}$
- 8: **end if**
- 9: **end for**
- 10: **end for**
- 11: Generate prediction surface:
- 12: **for** each point $(m, n) \notin (i, j)$ **do**
- 13: Compute $\hat{p}_{t+d}^{m,n} \leftarrow w_k \sum_{i,j=1}^N * \hat{p}_{t+d}^{i,j}$
- 14: **end for**
- 15: **return** pollution prediction surface on road network

longer path than the shortest path but optimize both distance and pollution levels for safe travel. The existing navigation system usually uses time and distance as the weight to compute the shortest path. Few of the existing models compute the least air polluted path based on the polluted area [25–29, 49] for the current time only, not for the different temporal resolution. Hence, to overcome this limitation, the proposed framework uses a deep learning approach to predict the safest path over the $t+d$ days, which will be the least exposure to air pollution. The pseudo-code of the proposed routing solution is presented in Algorithm 2.

6. Results and discussions

6.1. Parameter optimization

To demonstrate the proposed model and algorithm efficiency, four groups of models were experimented with using PM_{10} concentration data, which consists of a series of day-wise sampling values. After the preprocessing step, the model was trained where 80% of data was utilized as the training set (January 2005–October 2015), 10% (November 2015) as the validation remaining 10%, i.e., last month data of 2015 as a test set. We used two-layer of CNN, two-layer of LSTM with 1078 iterations, and one layer of the, having 343

Algorithm 2 Predicting the least air polluted route

Input: Road network graph $G \hat{=} (V, E, w)$, source s , destination h , starting time.

Output: Optimized route, directions, distance, duration

Initialization: Take $d(s)=0$ and $d(h)=\text{infinity}$

- 1: Generate interpolation prediction map:
- 2: Interpolation map with class level $c \leftarrow$ Call Algorithm 1
- 3: **for** each node h in Graph G **do**
- 4: **if** $h \neq s$ **then**
- 5: $d[h]=\text{infinity}$
- 6: push h in Queue Q
- 7: **else**
- 8: $c_h * \min d(h) \leftarrow \text{cost}(h)$
- 9: pop h node from Q
- 10: **end if**
- 11: **end for**
- 12: **for** each neighborhood point z of h **do**
- 13: $c_h \cdot \min d(h) + c_z \cdot \min d(z) \leftarrow \cos t(h, z)$
- 14: $\cos t(h, z) + d(h) \leftarrow \text{final cost}$
- 15: return final cost
- 16: **end for**

iterations for getting the best prediction value during temporal air quality modeling at the large temporal resolution to identify the most polluted location of Odisha over the next four weeks of December 2015. Furthermore, the learning rate set to 1e-3 and dropout is 0.2; Adam is used as an optimizer and Mean Square Error (MSE) as a loss function to train the model. The window size set to 50 and n-time steps is 28 to evaluate the model's long-term prediction performance.

6.2. Prediction performance comparison at large temporal resolution

Most of the existing air pollution prediction models predict air quality concentration levels for the next few hours for existing air pollution monitoring sites. Predicting air quality levels for the entire location can add profit to get optimum air quality prediction results for a considerable period. Generally, air quality prediction for an extended period has lower prediction accuracy than for a short duration. This issue might occur because of the small sample utilization for the long-term forecast. Hence, developing an air quality prediction model is required to perform air pollution prediction for the entire location and its road network for an extended period.

The first part of the proposed model has a neural network layer to perform air pollution prediction for all available monitoring stations only. Another vital property of the proposed method is that it has an EBK interpolation layer. The CLE model's execution

provides prediction value at each point, where no monitoring stations are available. This is the most important feature of the proposed method. It can be utilized by researchers and policymakers to get the overall air quality level of a particular location and its road network at a different time. Though this approach can give vital information for public safety, very few models can predict air quality levels at unmeasured locations. By conducting temporal-spatial interpolation of PM_{10} concentrations, it is possible to get a average PM_{10} distribution surface at large time granularity. The cross-validation technique evaluates the temporal-spatial prediction performance of the CLE model. During cross-validation, the forecast value is predicted at an observed location by considering the other observations and removing the original observed forecast value. Then the observed and temporal-spatial prediction value is compared to calculate the prediction error. The model performance was evaluated using error metrics like Root Mean Square Error (RMSE) [50]. It can be represented as below, [51, 52].

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2} \quad (12)$$

where y_t is the observed pollutant value at time t , \hat{y}_t presents predicted value at t and n is the total number of samples. The comparison table is shown in Table 1.

Table 1. Prediction comparison at high temporal resolution.

Method type	Model	RMSE
Deep learning	BILSTM [53]	79.53
Deep learning	BIGRU [54]	71.19
Deep learning	CNN-BILSTM [55]	65.20
Deep learning	CNN-BIGRU [50]	64.09
Deep learning	CNN-GRU [56]	60.45
Deep learning	LSTM [57–59]	47.84
Deep learning	GRU [60] [61]	43.06
Geostatistics	Exponential kriging [62]	25.60
Geostatistics	Universal kriging [62]	25.60
Deterministic	Inverse distance weighting [62]	24.07
Geostatistics	Spherical kriging [63]	23.67
Machine learning	Radial basis function [38]	23.03
Proposed model	CLE	18.16

The proposed CLE model is compared with baseline deep learning models to evaluate the model prediction efficiency for the next four weeks. It shows that the BILSTM model had the worst performance having the highest value of error metrics. Error metrics for BIGRU, CNN-BILSTM, CNN-BIGRU, CNN-GRU, LSTM, and GRU have reduced relatively. The comparative results show that the proposed model outperforms the other deep learning models in the case of long-term prediction due to the lower error metrics,

as the lower the error metrics value signifies the higher prediction performance. The proposed model's temporal-spatial prediction performance is compared with the deterministic model, e.g. (IDW), machine learning model, e.g., RBF neural network, geostatistical techniques, e.g., spherical, exponential, and universal kriging as represented in Table 1. It can be seen from the comparison table that the CLE model has lower error metrics values than any other geostatistical, deterministic, and machine learning-based spatial prediction model due to its time and space dimension integration capability. This result proves that it is necessary to take both space and time dimensions simultaneously instead of only performing spatial interpolation to better interpolation results.

Figure 4-5 shows the calculated PM_{10} prediction map by CLE model at weekly and two day granularities. As shown in Figure 4-5, highly polluted areas located nearer to the eastern ghats and eastern coastal plains of Odisha. As the east part of Odisha has the most developed part of the state, we can see a high level of pollution due to high traffic emissions. So the government has always taken the necessary steps to improve these locations' air quality than any other province. It also noticed that the average PM_{10} concentration ranges from $35 \mu\text{g}/\text{m}^3$ to $121 \mu\text{g}/\text{m}^3$, which shows the unfavorable exposure level of PM_{10} in Odisha.

6.3. Web Client to Trace temporal-spatial prediction of PM_{10}

The interpolated map is overlapped with the Google map using the ArcGIS [64] Online cloud server to design a prediction map web application. It is developed for visualization purposes, represented in Figure 6; it displays area size with its corresponding predicted class, minimum, and maximum pollution level. Web App Builder for ArcGIS service is utilized to build the web client application, and it has all facilities to create HTML/JAVA script featured web application. It also provides facilities to run the application online on its server.

6.4. Routing service to avoid polluted areas

When the web map layer displaying the prediction map is published, the computed polluted areas can be treated as input barriers while accessing a city's routes. ArcGIS Online network analysis feature is utilized to perform this operation. ArcGIS online routing service models driving and walking transportation modes. It provides facilities to compute routes based on the shortest distance and time from source to destination. It also allows us to consider traffic flow directions and restrictions. The network service routing layer enables treating the highly polluted area as input barriers;

it prohibits all the intersecting paths from routing analysis and predicts the least air pollution exposure paths.

Previous research was conducted to compute the pollution-free route by performing network analysis over IDW [25–28] interpolated map. IDW interpolation can predict average, maximum, and minimum air pollution concentration for the current and past time only, not for the future. It can also be seen that the prediction performance of IDW (RMSE=24.07) is less as compared to the CLE model (RMSE=18.16) due to high prediction error, as shown in Table 1.

This research work performed network analysis over the temporal-spatial interpolation prediction map instead of spatial prediction map using ArcGIS online to minimize air pollution level and travel time simultaneously. Therefore, it can provide an optimized routing solution at a different temporal resolution, unique from any other route navigation system. ArcGIS, a geographical information system [65], is utilized to visualize routing solutions. An IoT-based platform helps to access real-time traffic data with each route detail and overlaps the air pollution interpolation map with the street-based map. There may be multiple paths for a given source and destination, and the user may select any of them based upon their preferences. User preference of start and endpoint can be updated with the help of this IoT-based framework, and routing solutions can be accessible by the web application. To identify the proposed model's effectiveness, we evaluated its performance in terms of minimum pollution level and distance. Hence, we simulated the model by selecting 16 monitoring sites of Odisha and proposed the routing solution. The designed web application would help users make decisions more straightforward, as it can show both the shortest and the least polluted path and differentiate them.

The result of the optimized routing solution is compared with google map provided the shortest path, as shown in Figure 7. Figure 7 (a1) predicts the shortest path from the source to the destination point but passes through the highly polluted area, marked in red. It takes one hour 38 minutes for car drivers to cover the source to the destination point. At the same time, Figure 7 (b1) predicts the path for the next day (1st December 2015) for car drivers; that result is obtained from the proposed routing solution. It will take more time, i.e., 2 hours six minutes, to cover the source to the destination but pass through the least polluted area and predict a healthier path for traveling. Figure 8 predicts the shortest route and the least air polluted path for 2nd December 2015 for pedestrians. Figure 8 (c1) predicts the shortest path from Sophia junior college to Kendriya Vidyalaya, which needs 18 minutes to cross that path. In contrast, Figure 8 (d1) predicts the least polluted path, which will take four

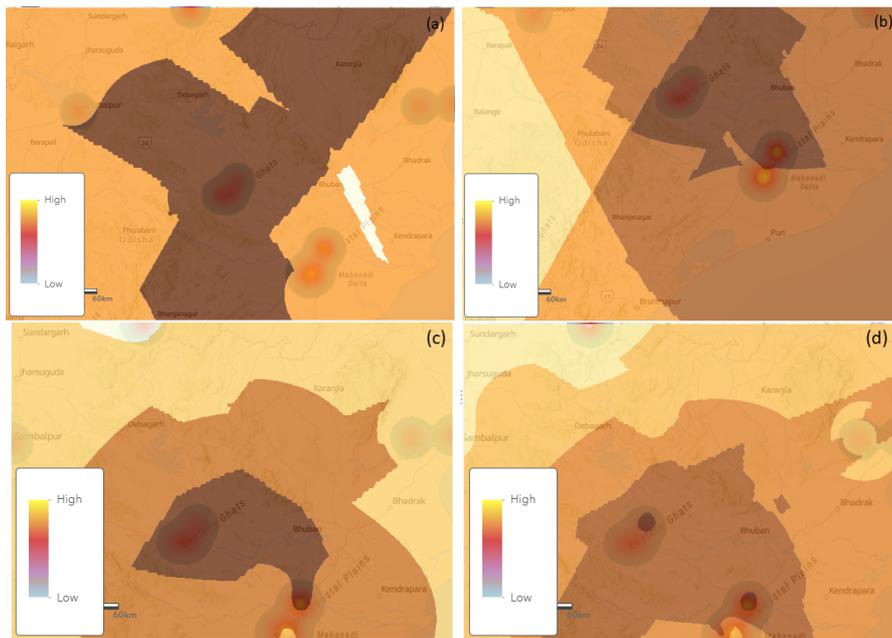


Figure 4. Spatial distribution of average PM_{10} concentration weekly in 2015 December. Figure (a), (b),(c),(d) predict the first, second, third and fourth week average PM_{10} concentrations. The color scale shows the pollutant variations over the study area.

minutes more for a pedestrian to cross it. It shows the positive side of the proposed algorithm. The designed web application predicts air pollution level and also recommend pollution-free paths for Odisha. Based on the user preference, they can choose the path for their safe journey.

7. Conclusion

For the first time, the present study proposes a newly designed deep learning-based temporal-spatial interpolation method to provide an optimized routing solution for users that minimizes air pollution exposure risk. The current work demonstrates models that take PM_{10} pollutant value into account to predict the healthier path in advance. The predicted least polluted path might be longer than the shortest path, but it avoids the pollution level that varies over space and time. The proposed method can be implemented in smart cities to a better quality of life by utilizing the temporal-spatial interpolation prediction map and proposed routing solution in advance. A comparative analysis is conducted, which shows the CLE model's effectiveness to recommend a healthier path for the next coming days.

The main finding of this research paper are briefly described as follows:

- This paper integrated machine learning, deep learning, and spatial prediction techniques for air quality modeling at high temporal-spatial

resolution. This could be essential information for public safety.

- The proposed model also provides effectively smaller prediction error at high temporal resolution than the baseline neural network models. It has almost 78%, 61% better air quality modeling performance than the ordinary BILSTM and GRU model, respectively.
- CLE model has a minimum of 21-29% better prediction performance at high temporal resolution than the traditional interpolation techniques. It can also solve the data imputation of geospatial air pollution data while conducting air quality prediction.
- This research paper utilizes the output of the proposed temporal-spatial interpolation model as input for GIS modeling, which prevents the highly polluted area as barriers to predict the healthier path at larger temporal resolution. So users can get overall air quality information on roads for the next four weeks.
- To the best of our knowledge, this is the first research study conducted in Odisha, India to predict air pollution levels and the safest path for the entire study area.

Still, this research work can further be improved. Meteorological factors, climate change, traffic emission, and rainfall have a direct and indirect impact on the

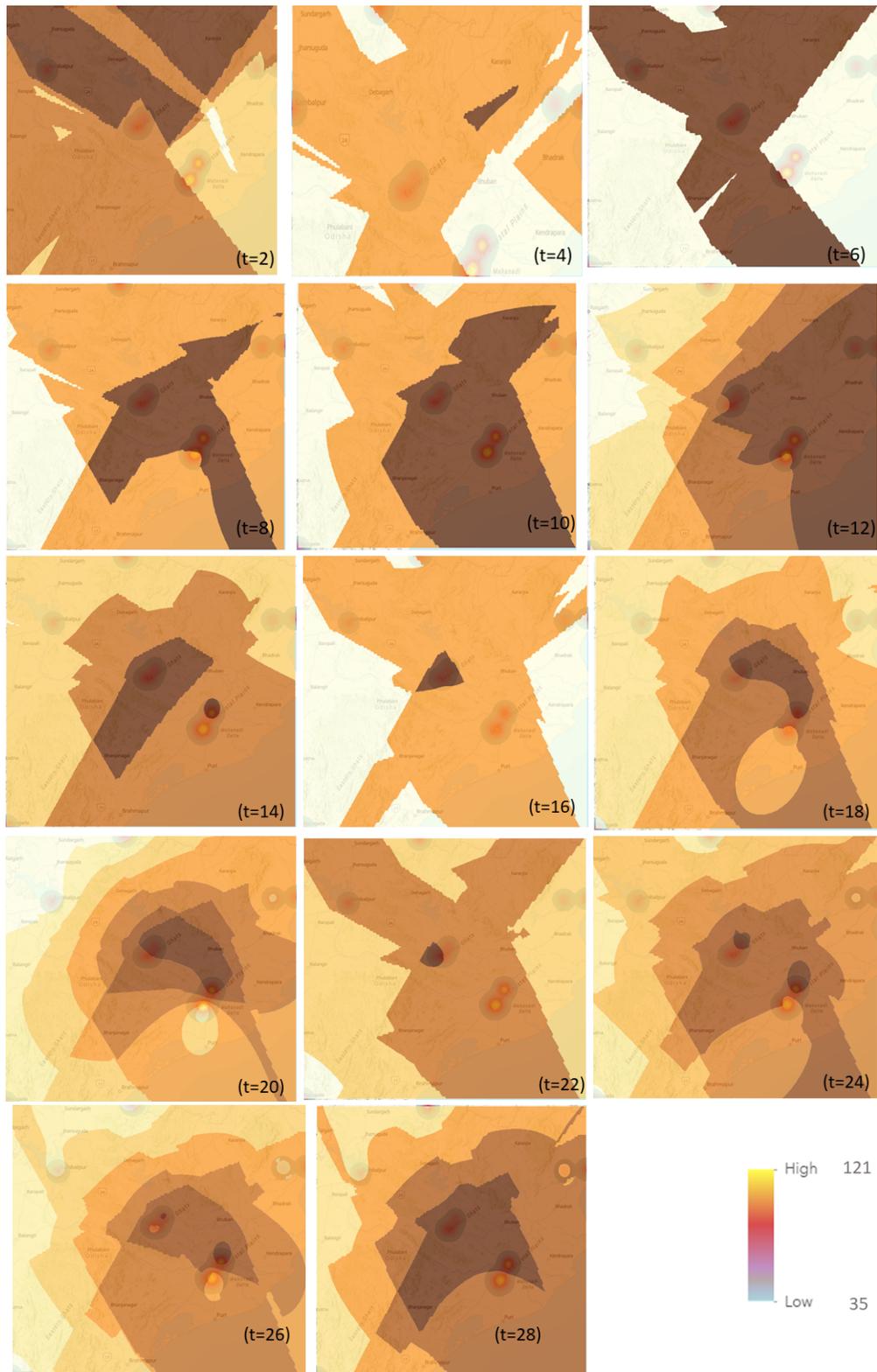


Figure 5. Spatial distribution of average PM_{10} value on every two days interval basis in December 2015.

ambient air pollution level, which should be analyzed during air quality modeling. We have not considered

the meteorological and traffic impact on air pollution due to lack of data availability.



Figure 6. Developed web application to display the air pollution prediction map of Odisha. The application shows the maximum and minimum pollution levels of a particular area with location area size.

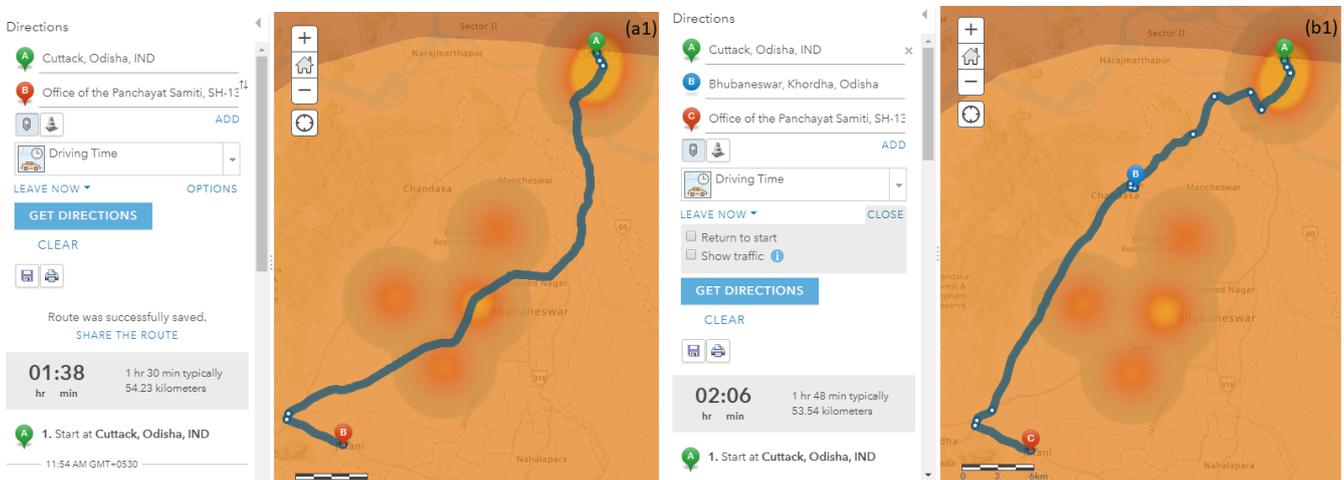


Figure 7. User interface (a1) predicts the highly polluted path, (b1) predicts the least polluted healthier path for the scheduled interval (01.12.2015) for a car driver.

There is a further scope of using this technology by implementing this concept on a large area rather than limiting it to a particular confined space. Along with that, more parameters can be included, which will be much more beneficial to the users by considering the drastic impact of those pollutants on individuals.

8. Acknowledgment

This research and product development work is financially supported by the Ministry of Human Resource Development and Ministry of Housing and Urban Affairs, Government of India [Grant No.-7794/2016].

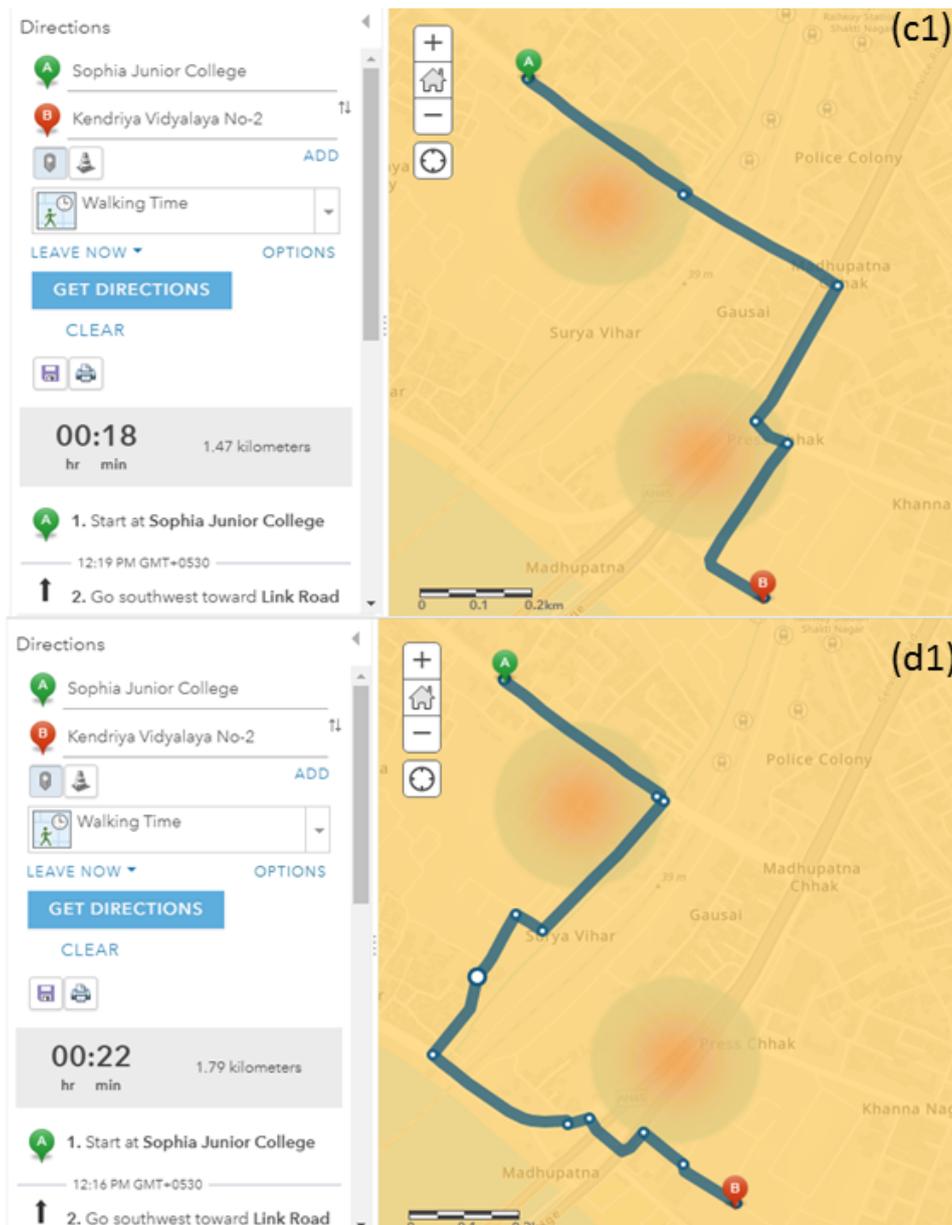


Figure 8. User interface (c1) predicts the highly polluted route, (d1) predicts the least polluted healthier path for the scheduled interval (02.12.2015) for a pedestrian.

References

- [1] SAMAL, K.K.R., BABU, K.S. and DAS, S.K. (2021) Multi-directional temporal convolutional artificial neural network for pm2. 5 forecasting with missing values: A deep learning approach. *Urban Climate* **36**: 100800.
- [2] SAMAL, K.K.R., BABU, K.S. and DAS, S.K. (2021) Temporal convolutional denoising autoencoder network for air pollution prediction with missing values. *Urban Climate* **38**: 100872.
- [3] BI, J., STOWELL, J., SETO, E.Y., ENGLISH, P.B., AL-HAMDAN, M.Z., KINNEY, P.L., FREEDMAN, F.R. *et al.* (2020) Contribution of low-cost sensor measurements to the prediction of pm2. 5 levels: A case study in imperial county, california, usa. *Environmental research* **180**: 108810.
- [4] MAHAJAN, S., TANG, Y.S., WU, D.Y., TSAI, T.C. and CHEN, L.J. (2019) Car: The clean air routing algorithm for path navigation with minimal pm2. 5 exposure on the move. *IEEE Access* **7**: 147373–147382.
- [5] SINGH, R., ZHANG, Y., WANG, H., MIAO, Y. and AHMED, K. (2020) Investigation of social behaviour patterns using location-based data-a melbourne case study. *EAI Endorsed Transactions on Scalable Information Systems*.
- [6] AWAD, Y.A., KOUTRAKIS, P., COULL, B.A. and SCHWARTZ, J. (2017) A spatio-temporal prediction model based on support vector machine regression: Ambient black carbon in three new england states. *Environmental*

- research **159**: 427–434.
- [7] CORANI, G. and SCANAGATTA, M. (2016) Air pollution prediction via multi-label classification. *Environmental Modelling & Software* **80**: 259–264.
 - [8] JOHANSSON, L., EPITROPOU, V., KARATZAS, K., KARPPINEN, A., WANNER, L., VROCHIDIS, S., BASSOUKOS, A. *et al.* (2015) Fusion of meteorological and air quality data extracted from the web for personalized environmental information services. *Environmental Modelling & Software* **64**: 143–155.
 - [9] SOLAZZO, E., CAI, X. and VARDOULAKIS, S. (2009) Improved parameterisation for the numerical modelling of air pollution within an urban street canyon. *Environmental Modelling & Software* **24**(3): 381–388.
 - [10] CABANEROS, S.M.S., CALAUTIT, J.K. and HUGHES, B.R. (2019) A review of artificial neural network models for ambient air pollution prediction. *Environmental Modelling & Software*.
 - [11] ZHANG, B., ZHANG, H., ZHAO, G. and LIAN, J. (2020) Constructing a pm2.5 concentration prediction model by combining auto-encoder with bi-lstm neural networks. *Environmental Modelling & Software* **124**: 104600.
 - [12] SHAHZADA, A. and ASKAR, K. (2011) Dynamic vehicle navigation: An a* algorithm based approach using traffic and road information. In *2011 IEEE International Conference on Computer Applications and Industrial Electronics (ICCAIE)* (IEEE): 514–518.
 - [13] AKAY, A.E., WING, M.G., SIVRIKAYA, F. and SAKAR, D. (2012) A gis-based decision support system for determining the shortest and safest route to forest fires: a case study in mediterranean region of turkey. *Environmental monitoring and assessment* **184**(3): 1391–1407.
 - [14] ABOUSAEIDI, M., FAUZI, R. and MUHAMAD, R. (2016) Geographic information system (gis) modeling approach to determine the fastest delivery routes. *Saudi journal of biological sciences* **23**(5): 555–564.
 - [15] PRAPULLA, S., RAO, S.N. and HERUR, V.A. (2017) Road quality analysis and mapping for faster and safer travel. In *2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS)* (IEEE): 2487–2490.
 - [16] GALBRUN, E., PELECHRINIS, K. and TERZI, E. (2016) Urban navigation beyond shortest route: The case of safe paths. *Information Systems* **57**: 160–171.
 - [17] RUK, S.A., KHAN, S.G., SALAHUDDIN SADDAR, N.A.M., SHAH, S.M.Z. and BROHI, S. (2018) Evaluating route security and predicting the safest alternative using risk factors. *Indian Journal of Science and Technology* **11**: 27.
 - [18] KELER, A. and MAZIMPAKA, J.D. (2016) Safety-aware routing for motorised tourists based on open data and vgi. *Journal of location Based services* **10**(1): 64–77.
 - [19] RAJU, P.M.D.S.G. *et al.* (2013) Identification of accident black spots for national highway using gis. *International Journal of Scientific & Technology Research* **2**(2): 154–157.
 - [20] ZHOU, E., MAO, S. and LI, M. (2017) Investigating street accident characteristics and optimal safe route recommendation: A case study of new york city. In *2017 25th International Conference on Geoinformatics (IEEE)*: 1–7.
 - [21] COSTABILE, F. and ALLEGRINI, I. (2008) A new approach to link transport emissions and air quality: An intelligent transport system based on the control of traffic air pollution. *Environmental Modelling & Software* **23**(3): 258–267.
 - [22] TOMÁS, V.R., PLA-CASTELLS, M., MARTÍNEZ, J.J. and MARTÍNEZ, J. (2016) Forecasting adverse weather situations in the road network. *IEEE Transactions on Intelligent Transportation Systems* **17**(8): 2334–2343.
 - [23] KIESEWETTER, G., SCHOEPP, W., HEYES, C. and AMANN, M. (2015) Modelling pm2.5 impact indicators in europe: health effects and legal compliance. *Environmental Modelling & Software* **74**: 201–211.
 - [24] MÜLLER, S. and VOISARD, A. (2015) Air quality adjusted routing for cyclists and pedestrians. In *Proceedings of the 1st ACM SIGSPATIAL International Workshop on the Use of GIS in Emergency Management (ACM)*: 19.
 - [25] RAMOS, F., TRILLES, S., MUÑOZ, A. and HUERTA, J. (2018) Promoting pollution-free routes in smart cities using air quality sensor networks. *Sensors* **18**(8): 2507.
 - [26] SHARKER, M.H. and KARIMI, H.A. (2014) Computing least air pollution exposure routes. *International Journal of Geographical Information Science* **28**(2): 343–362.
 - [27] VAMSHI, B. and PRASAD, R.V. (2018) Dynamic route planning framework for minimal air pollution exposure in urban road transportation systems. In *2018 IEEE 4th World Forum on Internet of Things (WF-IoT)* (IEEE): 540–545.
 - [28] LI, H.C., CHIUH, P.T., LIU, S.P. and HUANG, Y.Y. (2017) Assessment of different route choice on commuters' exposure to air pollution in taipei, taiwan. *Environmental Science and Pollution Research* **24**(3): 3163–3171.
 - [29] ZAHMATKESH, H., SABER, M. and MALEKPOUR, M. (2015) A new method for urban travel route planning based on air pollution sensor data. *Curr World Environ* **10**: 699–704.
 - [30] BUTEAU, S., HATZOPOULOU, M., CROUSE, D.L., SMARGIASSI, A., BURNETT, R.T., LOGAN, T., CAVELLIN, L.D. *et al.* (2017) Comparison of spatiotemporal prediction models of daily exposure of individuals to ambient nitrogen dioxide and ozone in montreal, canada. *Environmental research* **156**: 201–230.
 - [31] SAMAL, K., BABU, K.S., DAS, S.K. and ACHARAYA, A. (2019) Time series based air pollution forecasting using sarima and prophet model. In *Proceedings of the 2019 International Conference on Information Technology and Computer Communications (ACM)*: 80–85. doi:<https://doi.org/10.1145/3355402.3355417>.
 - [32] CENTRALPOLLUTIONCONTROLBOARD (2018, Jan 6), Air pollution. URL <http://cpcb.nic.in/air-pollution/>.
 - [33] SAMAL, K.K.R., BABU, K.S. and DAS, S.K. (2020) Ors: The optimal routing solution for smart city users. In *Electronic Systems and Intelligent Computing* (Springer), 177–186.
 - [34] SAMAL, K., BABU, K.S. and DAS, S.K. (2021) Spatio-temporal prediction of air quality using distance based interpolation and deep learning techniques. *EAI Endorsed Transactions on Smart Cities* **5**(14): e4.
 - [35] SAMAL, K.K.R., BABU, K.S., ACHARYA, A. and DAS, S.K. (2020) Long term forecasting of ambient air quality using deep learning approach. In *2020 IEEE 17th India*

- Council International Conference (INDICON)* (IEEE): 1–6.
- [36] SAMAL, K.K.R., BABU, K.S., PANDA, A.K. and DAS, S.K. (2020) Data driven multivariate air quality forecasting using dynamic fine tuning autoencoder layer. In *2020 IEEE 17th India Council International Conference (INDICON)* (IEEE): 1–6.
- [37] SAMAL, K.K.R., PANDA, A.K., BABU, K.S. and DAS, S.K. (2021) An improved pollution forecasting model with meteorological impact using multiple imputation and fine-tuning approach. *Sustainable Cities and Society* **70**: 102923.
- [38] BHUNIA, G.S., SHIT, P.K. and MAITI, R. (2018) Comparison of gis-based interpolation methods for spatial distribution of soil organic carbon (soc). *Journal of the Saudi Society of Agricultural Sciences* **17**(2): 114–126.
- [39] YANG, Z. and WANG, J. (2017) A new air quality monitoring and early warning system: Air quality assessment and air pollutant concentration prediction. *Environmental research* **158**: 105–117.
- [40] FENG, X., LI, Q., ZHU, Y., HOU, J. and WANG, J. (2015) An estimate of population exposure to automobile source pm 2.5 in beijing using spatiotemporal analysis. In *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)* (IEEE): 3029–3032.
- [41] ELUMALAI, V., BRINDHA, K., SITHOLE, B. and LAKSHMANAN, E. (2017) Spatial interpolation methods and geostatistics for mapping groundwater contamination in a coastal area. *Environmental Science and Pollution Research* **24**(12): 11601–11617.
- [42] VORAPRACHA, P., PHONPRASERT, P., KHANARUKSOMBAT, S. and PIJARN, N. (2015) A comparison of spatial interpolation methods for predicting concentrations of particle pollution (pm10). *International Journal of Chemical, Environmental and Biological Sciences* **3**(4): 302–306.
- [43] CONTRERAS-OCHANDO, L. and FERRI, C. (2016) airvlc: An application for visualizing wind-sensitive interpolation of urban air pollution forecasts. In *2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW)* (IEEE): 1296–1299.
- [44] GORAI, A.K., TCHOUNWOU, P.B. and MITRA, G. (2017) Spatial variation of ground level ozone concentrations and its health impacts in an urban area in india. *Aerosol and air quality research* **17**(4): 951.
- [45] NEBENZAL, A., FISHBAIN, B. and KENDLER, S. (2020) Model-based dense air pollution maps from sparse sensing in multi-source scenarios. *Environmental Modelling & Software* : 104701.
- [46] CURTARELLI, M., LEÃO, J., OGASHAWARA, I., LORENZZETTI, J. and STECH, J. (2015) Assessment of spatial interpolation methods to map the bathymetry of an amazonian hydroelectric reservoir to aid in decision making for water management. *ISPRS International Journal of Geo-Information* **4**(1): 220–235.
- [47] ARSENI, M., VOICULESCU, M., GEORGESCU, L.P., ITICESCU, C. and ROSU, A. (2019) Testing different interpolation methods based on single beam echosounder river surveying. case study: Siret river. *ISPRS International Journal of Geo-Information* **8**(11): 507.
- [48] KRIVORUCHKO, K. and GRIBOV, A. (2019) Evaluation of empirical bayesian kriging. *Spatial Statistics* **32**: 100368.
- [49] BRIGGS, D.J., COLLINS, S., ELLIOTT, P., FISCHER, P., KINGHAM, S., LEBRET, E., PRYL, K. *et al.* (1997) Mapping urban air pollution using gis: a regression-based approach. *International Journal of Geographical Information Science* **11**(7): 699–718.
- [50] TAO, Q., LIU, F., LI, Y. and SIDOROV, D. (2019) Air pollution forecasting using a deep learning model based on 1d convnets and bidirectional gru. *IEEE Access* **7**: 76690–76698.
- [51] LEE, M.H., RAHMAN, N.H.A., LATIF, M.T., NOR, M.E., KAMISAN, N.A.B. *et al.* (2012) Seasonal arima for forecasting air pollution index: A case study. *American Journal of Applied Sciences* **9**(4): 570–578.
- [52] ZOU, B., LI, S., ZHENG, Z., ZHAN, B.F., YANG, Z. and WAN, N. (2020) Healthier routes planning: A new method and online implementation for minimizing air pollution exposure risk. *Computers, Environment and Urban Systems* **80**: 101456.
- [53] MADAAN, D., DUA, R., MUKHERJEE, P. and LALL, B. (2019) Vayuanukulani: Adaptive memory networks for air pollution forecasting. *arXiv preprint arXiv:1904.03977* .
- [54] QIN, H. (2019) Comparison of deep learning models on time series forecasting: a case study of dissolved oxygen prediction. *arXiv preprint arXiv:1911.08414* .
- [55] DU, S., LI, T., YANG, Y. and HORNG, S.J. (2018) Deep air quality forecasting using hybrid deep learning framework. *arXiv preprint arXiv:1812.04783* .
- [56] DU, S., LI, T., GONG, X., YU, Z., HUANG, Y. and HORNG, S.J. (2018) A hybrid method for traffic flow forecasting using multimodal deep learning. *arXiv preprint arXiv:1803.02099* .
- [57] MA, J., DING, Y., CHENG, J.C., JIANG, F. and WAN, Z. (2019) A temporal-spatial interpolation and extrapolation method based on geographic long short-term memory neural network for pm2. 5. *Journal of Cleaner Production* **237**: 117729.
- [58] KÖK, İ., ŞİMŞEK, M.U. and ÖZDEMİR, S. (2017) A deep learning model for air quality prediction in smart cities. In *2017 IEEE International Conference on Big Data (Big Data)* (IEEE): 1983–1990.
- [59] SONG, X., LIU, Y., XUE, L., WANG, J., ZHANG, J., WANG, J., JIANG, L. *et al.* (2019) Time-series well performance prediction based on long short-term memory (lstm) neural network model. *Journal of Petroleum Science and Engineering* : 106682.
- [60] ATHIRA, V., GEETHA, P., VINAYAKUMAR, R. and SOMAN, K. (2018) Deepairnet: Applying recurrent networks for air quality prediction. *Procedia computer science* **132**: 1394–1403.
- [61] ZHOU, X., XU, J., ZENG, P. and MENG, X. (2019) Air pollutant concentration prediction based on gru method. In *Journal of Physics: Conference Series* (IOP Publishing), **1168**: 032058.
- [62] SON, J.Y., BELL, M.L. and LEE, J.T. (2010) Individual exposure to air pollution and lung function in korea: spatial analysis using multiple exposure approaches. *Environmental research* **110**(8): 739–749.
- [63] GONG, G., MATTEVADA, S. and O'BRYANT, S.E. (2014) Comparison of the accuracy of kriging and idw interpolations in estimating groundwater arsenic concentrations in texas. *Environmental research* **130**: 59–69.

- [64] BOOTH, B., MITCHELL, A. *et al.* (2001), Getting started with arcgis. for air pollution exposure assessment. *Environmental Modelling & Software* **105**: 17–23.
- [65] MORLEY, D.W. and GULLIVER, J. (2018) A land use regression variable generation, modelling and prediction tool