EAI Endorsed Transactions
on Smart Cities

Research Article EAI.EU

Spatio-temporal Prediction of Air Quality using Distance Based Interpolation and Deep Learning Techniques

K. Krishna Rani Samal[1,*], Korra Sathya Babu[1], Santos Kumar Das[1]

[1]National Institute of Technology, Rourkela, India

Abstract

The harmful impact of air pollution has drawn raising concerns from ordinary citizens, researchers, policymakers, and smart city users. It is of great importance to identify air pollution levels at the spatial resolution on time so that its negative impact on human health and environment can be minimized. This paper proposed the CNN-BILSTM-IDW model, which aims to predict and spatially analyze the pollutant level in the study area in advance using past observations. The neural network-based Convolutional Bidirectional Long short-term memory (CNN-BILSTM) network is employed to perform time series prediction over the next four weeks. Inverse Distance Weighting (IDW) is utilized to perform spatial prediction. The proposed CNN-BILSTM-IDW model provides almost 16% better prediction performance than the ordinary IDW method, which fails to predict spatial prediction at a high temporal period. The results of the presented comparative analysis signify the efficiency of the proposed model.

Received on 20 June 2020; accepted on 29 November 2020; published on 15 January 2021

Keywords: Air quality, Deep Learning, LSTM, Inverse Distance Weighting, Spatio-temporal prediction

doi:10.4108/eai.15-1-2021.168139

1. Introduction

Air pollution has become a severe problem for many developing countries in the world. India is one of them (Brauer et al. 2019; Pant, Guttikunda, and Peltier 2016). With the rapid growth of urbanization, global consumption of fossil fuels, and oil, air pollution can cause significant health issues and affect human body parts severely. High exposure to air pollutants and other gases can cause a severe asthma attack and many more diseases. Due to the poor quality of atmospheric air, people are more vulnerable to suffering from asthma, lung cancer, and respiratory infections. Commonly seen air pollutants such as PM_{10}, $PM_{2.5}$, SO_2, NO_2 and O_3 are more responsible for heart attack, lung diseases, and respiratory problems. Million of people are dying worldwide every year due to this type of disease. Air quality in many Indian cities failed to obey many international and national standards and CPCB (Beig et al. 2020) effective pollution control strategies. According to the NCAP report, 43 smart cities of India are falling under 102 nonattainment cities of the country. More than half of the country's population is exposed to particle matter, which exceeds the permissible limits. Recent global air pollution research studies say that almost 600000 premature death per year occurs in India due to ambient air pollution level (Hama et al. 2020). Air pollution control in India has become challenging due to the high impact of meteorological factors and traffic emission (Sharma, Kharol, and Badarinath 2010), which is very difficult to analyze.

Among all the polluted cities in India, Odisha has six nonattainment cities. It has been observed seriously that many people of Odisha are suffering from chronic diseases due to reduced air quality levels as it has become one of the polluted states in India. From a descriptive statistical analysis of health care data of this state, it is found that 11951 the number of females and 8454 males per 10000 population of Khordha district affected by acute illness during 2012-2013 caused by air pollution. Moreover, Puri, Jagatsinghapur, Khorda,

*Corresponding author. Email: 517cs6019@nitrklac.in

Nayagarh, Cuttack district people are mostly affected by asthma disease, whose primary source is environmental air pollution (Samal, Babu, Santosh Kumar Das, et al. 2019). Therefore, it became necessary to predict the air pollution boundaries and its spatial distribution to regulate it.

Concerning the severe negative impact of air pollution, the government has taken several smart initiatives, also working with different research institutions to take essential steps against ambient air pollution levels. The government has also developed many air pollution monitoring stations to collect air quality data, which can be utilized to forecast air pollution levels for the next hour, day, or week. The forecasting result provides timely information to take necessary prevention in advance. Thus air quality modeling and monitoring can help to mitigate the impact of air pollution. Several techniques have implemented to predict air pollution, i.e., deterministic, statistical, machine learning, and deep learning models.

These are the widely used techniques for air quality prediction. Commonly used deterministic methods are Weather Research and Forecasting models (WRF) (Saide et al. 2011), Nested Air Quality Prediction Modeling System (NAQPMS) (Z. Wang et al. 2001). Getting prediction results of these methods are expensive. These methods utilize the default parameters, so prediction results are also inappropriate in real-time scenarios. Statistical models are another kind of prediction model which overcomes the limitation of deterministic methods by utilizing a large amount of observed dataset. Statistical models such as ARIMA (Yenidoğan et al. 2018), SARIMA (Samal, Babu, Santosh Kumar Das, et al. 2019; Voynikova et al. 2015; M. H. Lee et al. 2012; N.-U. Lee et al. 2018), General Additive Models (GAMS), Geographically Weighted Regression and Multi-layer Regression (MLR) (McKendry 2002) have been utilized in air quality prediction. These statistical models are based on data stationarity and data linearity. These models assume the linear relationship between the observed and predicted value and incapable of handling data nonstationarity. So these statistical models have limited predicted performance.

To address these problems, researchers and policymakers adopted machine learning models such as Support Vector Machine, Random Forest (Zamani Joharestani et al. 2019), Artificial Neural Network (ANN) (Elangasinghe et al. 2014), Fuzzy Neural Network (Zhou, W. Li, and Qiao 2017; Zahedi et al. 2014), Linear Regression, and xgboost (Pan 2018; Zamani Joharestani et al. 2019). Feed forward neural network-based ANN has shown better air quality prediction performance. Though these methods have shown better accuracy in air quality prediction, these shallow neural network models fail to analyze the correlation among features of a multivariate air pollution dataset.

The time series pollution dataset has long term dependency among all features. With the rapid development of artificial intelligence techniques, machine learning models no longer remain as the state of the art models. Several researchers have conducted air quality modeling using deep learning techniques and have proven better prediction models than machine learning in terms of temporal analysis of the air pollution dataset. Deep learning models have shown better performance in sequential modeling, human detection, medical image classification, and many more applications. Deep learning models, i.e., Recurrent Neural Network (RNN), LSTM, Gated Recurrent Unit (GRU) (Du et al. 2019) models, have also played an essential role in air quality prediction. Few researchers added a Convolutional Neural Network (CNN) layer with the shallow, deep learning models to capture the spatial features in the available time-series dataset, which give better prediction performance by analyzing both the spatial and temporal characteristics.

Most of the existing prediction models predict air pollution levels for the next hours for a particular site. Predicting air pollution levels for the entire study area for a long term period can add an advantage to get better air pollution prediction results. Usually, air pollution prediction performance for a long term period gives lower accuracy than for the short term period. This might be due to the small number of samples to perform long term air quality prediction. Therefore, it is essential to develop air pollution prediction models that can effectively perform air pollution prediction for the entire study area at a more significant time resolution.

To address this limitation, the current research study developed a methodology framework that follows a deep learning-based CNN-BILSTM layer for feature analysis and time series prediction. On the top of the CNN-BILSTM layer, the distance-based Inverse Distance Weighting interpolation layer is developed to perform spatial prediction for the entire study area at a higher temporal resolution. So, predicting the spatial variability for the next few days will surely help to ensure public safety.

Following the introduction, the rest is organized as follows. Section 2 represented related work. Section 3 and Section 4 include problem statements and study areas, respectively. Section 5 and Section 6 describes the proposed methodologies framework and results of the implemented experiments. Section 7 consists of the conclusion part of this research paper.

2. Related works

It is unreasonable and cost-effective to establish air quality monitoring stations to analyze air pollution levels at each corner of the study area. So, it has become a crucial problem to identify the spatial distribution of

air quality level and predicts its value for the entire study area. To overcome this issue, the government has launched satellites to monitor the air pollution level for the entire area, which can also provide a full mapping of atmospheric air pollution levels (Boys et al. 2014; S. Chowdhury et al. 2019). Much research has been conducted to predict air pollution levels using satellite-derived remote sensing images. Though satellite can capture the overall air pollution level for the entire study area, it can not capture a particular location air pollution level all the time. The captured remote sensing images are also blurred in nature due to cloud influence and movable satellites. Therefore, identify the spatial distribution of air pollution levels could be limited due to satellite-derived air pollution data.

Considering this limitation of air quality monitoring using satellite data, few researchers experimented with statistical spatial prediction models to analyze air pollution levels spatially (Gulliver et al. 2011). The spatial prediction model handles missing values of air pollution data obtained due to unavailable monitoring stations in a particular location. The spatial prediction model includes several deterministic models like Inverse Distance Weighting (IDW) and geostatistics models like Ordinary Kriging (OK) (Feng et al. 2015; Contreras-Ochando and Ferri 2016), Simple Kriging (Cressie 1990), Universal Kriging (Vorapracha et al. 2015), and Empirical Bayesian Kriging (EBK) models (Gunarathna, Kumari, and Nirmanee 2016). These models are efficient enough to predict air pollution levels for each monitoring station and also for unmeasured locations (Cressie 1990) but these models have limited prediction performance due to default predefined parameter settings.

To overcome these types of limitations of spatial prediction models, recently, many research studies have adopted machine learning techniques for spatial prediction of air quality data, such as Radial Basis Function (RBF) (Zou et al. 2015) and Artificial Neural Network (ANN) (Nevtipilova et al. 2014). But their prediction accuracy could be limited due to lack of temporal analysis. Due to the absence of temporal parameter analysis, these models can not predict the spatial distribution of air pollution for different period i.e., for short term and long term periods. Though the discussed models are efficient enough to predict the spatial distribution of pollutants for the current time, not for the future, these are not so useful to make proper decisions and safety measures in advance. So, it has of great importance of temporal modeling for air quality prediction.

To analyze the temporal component of air pollution data set, several machine learning based uni-variate air quality prediction models adopted. Linear regression, Support Vector Machine (SVM) (Suykens and Vandewalle 1999; J. Wang, Niu, and R. Wang 2017; Shaban, Kadri, and Rezk 2016), Random Forest (RF) (Zhu et al. 2018), Decision Tree (DT) (Safavian and Landgrebe 1991), xgboost, Multi-layer Perceptron (MLP) are the mostly used univariate time series prediction models. Uni-variate prediction models failed to analyze the temporal and spatial components simultaneously. These univariate models also do not support correlation analysis among meteorological factors e.g., atmospheric temperature, rainfall, wind speed (WS), wind direction (WD), relative humidity (RH) and air pollution.

To get rid of this issue, deep learning techniques evolved with the increasing demand of artificial intelligence techniques. Initially, deep learning-based Artificial Neural Network (ANN) experimented for air quality prediction. ANN was originated in the 1970s. Usually, it has one input, one output, and multiple hidden states. In ANN, the weight calculation of input data is treated as neurons for the next layer. However, when it comes to handling time series data set, the ANN network is unable to handle longer sequence data, and can not relate the current and future data with historical data. To resolve these issues, research scholars developed a Recurrent Neural Network (RNN) (Fan et al. 2017), which is based on the ANN network. It takes the output of the first layer as input to the next layer, so it is able to transfer weight as a neuron. However, when it comes to dealing with more extended sequential data, it can not deal with them. As sequential time series air pollution data has a longer dependency on past observation due to effect meteorological factors. It is better to use a model that can deal with this type of issue. But RNN can not capture the long term dependency of long sequential air pollution data. It is also very challenging to train the RNN model due to gradient vanishing and exploding issues. To address these problems, Hochreiter proposed LSTM model (Fu, Z. Zhang, and L. Li 2016; B. Wang, Kong, and Guan 2019) in 1997, and Kyunghyun developed Gated Recurrent Unit (GRU) (Fu, Z. Zhang, and L. Li 2016; Tao et al. 2019) network to handle long term dependency in sequential time series data in 2014. Though the GRU model requires less parameter and less time to train, LSTM is proved as a more accurate prediction model for longer sequential data. It is also seen that a combination of multiple models has better prediction performance than shallow prediction models. Chiou-Jye et al. (Huang and Kuo 2018) presented a Spatio-temporal CNN-LSTM model to estimate air quality prediction level, which can capture both spatial and temporal features available in time series pollution dataset. It can also capture long term dependency on the air pollution dataset. It combined the wind speed, rainfall, and PM2.5 concentration information to train the model for air pollution prediction. Haofei Xie et al. (H. Xie et al. 2019) developed the CNN-GRU model , which can automatically extract the spatiotemporal

components of multidimensional and multi-station data. It also analyzes the impact of meteorological factors on air quality concentration levels for short term air pollution prediction.

Despite the memory capacity, the self-learning ability of neural networks, these models failed to capture the dependency of historical pollution level and the supporting information from nearby monitoring stations. Therefore, unable to analyze the temporal trend and spatial correlation simultaneously. Based on the above survey, the present work tried to predict air pollution levels at a high spatial and temporal resolution without using real-time sensors everywhere of the study area to mitigate the limitation of existing work and take necessary preventive action against the dangerous condition of the air pollution.

The main contribution of this research paper can be summarized as follows,

- Traditional interpolation techniques support spatial prediction for the present time. In contrast, the CNN-BILSTM-IDW model utilized past information to predict the spatial distribution of air quality for the future at better accuracy.
- The CNN-BILSTM-IDW model can effectively perform spatial prediction of PM_{10} level over a long period, i.e., for the next four weeks.
- The proposed spatial-temporal prediction method can effectively solve data imputation problems for air quality modeling by recovering missing attributes values.
- ArcGIS online is utilized to develop mobile and web applications to access timely information.

3. Problem Statement

Much of the pollution data available for different locations are sparse. Hence, there is a requirement for predicting continuous data from the available sparse dataset. In order to predict the interpolated surface with continuous pollution levels from variable data of different geolocations, the proper data analysis should be done with the efficient prediction model.

The mathematical formulation of spatial interpolation can be expressed as follows: Estimate the value of regionalized variable z_i, $\{z_i \in R^n | i = 1, 2, 3.....N\}$, at discrete points $m_i = \{(x_i^1, y_i^1), (x_i^2, y_i^2), ..(x_i^d, y_i^d)\}$, $\{m_i \in R | i = 1, 2, 3...N\}$ by considering N number of different existing neighborhood point values inside searching neighborhood area (r) with d-dimensional space. Weight (w) assignment will be done based on their distance from neighborhood points or autocorrelation among those points which signifies the influence of neighborhood points in estimation of a particular point value. Weight assignment is treated as function (f) such that,

$$f : R^n \rightarrow R$$
$$f(m_i) = z_i$$

4. Study area

Odisha, one of the polluted state of India (Tripathy and Dash 2018; Nayak and I. R. Chowdhury 2018), is selected as the study area for the research activity. It is reported that most of the industrial estate and smart cities of this state do not fulfill the air quality standard decided by the Central Pollution Control Board. The state is having very few numbers of continuous air pollution monitoring stations controlled by CPCB. Most of the sites are manual air quality monitoring stations that collect air quality data on the daily granularity level. The past observation data includes air quality dataset from 2005 to 2015. The data set contains PM_{10}, $PM_{2.5}$, SO_2, NO_2 air pollutant details with sampling date, and sampling location geographical attributes. It has seen from the analysis that PM_{10} pollutants have maximum air pollution contribution among all the contaminants in Odisha. So PM_{10} pollutant value is considered for experimental purposes in the study area. The figures are in one-millionth of a gram unit. The dataset includes 31 monitoring sites, but due to a large number of missing values, the research work considered only 16 monitoring sites for evaluation purposes.

5. Experimental Method

To validate the usefulness of the proposed methodology framework, ten years (01.01.2005-28.12.2015) dataset are collected from Odisha State Pollution Control Board (Odisha 2017, Oct 16; Samal, Babu, and Santos Kumar Das 2020; Samal, Babu, Santosh Kumar Das, et al. 2019). After data collection, data normalization and linear interpolation techniques are applied in the prepossessing step to get useful information for training the model. 90% of data are used for training purposes, whereas other remaining 10% are utilized for testing purposes. Adam and Mean Squared Error are implemented as optimizer and loss function respectively to train the proposed model. Each step of the proposed framework is presented in Figure 1. Each layer of the proposed methodology framework is discussed in the below subsections.

5.1. 1D ConvNet for feature learning

The 1D ConvNet (O'Shea and Nash 2015) is usually used for feature learning. It allows us to extract features from inputs and improve data efficiency by reducing data dimensionality. The weight sharing feature of 1D ConvNet minimizes the number of parameters of the multivariate time series dataset by increasing the

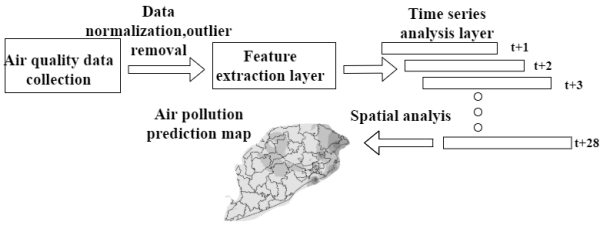

Figure 1. The proposed CNN-BILSTM-IDW architecture.

learning of the model. A learned pattern at a particular point of a sequence can be captured at other locations due to the same input transformation at each point. So, the change in the local trend of multivariate features can be determined. After completion of 1D ConvNet, the max-pooling operation is conducted to extract the maximum value of sub sequences of the dataset further to reduce the dimensionality of the input data source. The CNN layer performs three operations i.e., convolution, activation, and pooling respectively, which can be computed as below (Du et al. 2019),

$$c_n^l = \sum_m x_m^{l-1} * W_{mn}^l + b_n^l \tag{1}$$

$$x_n^l = \mathrm{Re}LU(c_n^l) \tag{2}$$

$$x_n^l = Flatten(x_n^l) \tag{3}$$

$$x_o^{l+1} = FCL(W_{on}^{l+1} x_n^l + b_o^{l+1}) \tag{4}$$

The convolution layer can be modeled using Equation 1-2 where $*$, W_{mn}^l, b_n^l represent convolution operator, filter and biases respectively. ReLU function is implemented as an activation function. x_m^{l-1} and c_n^l are the input, output vector to a convolution layer. The proposed architecture used two convolution layer for feature learning; l is used as the involved layer. The output of the preprocessing step is utilized as input to the CNN layer, where the learned representation of each segment is used as input to the next layer to model a hierarchical representation of features. After convolution operation, a flatten layer is added to transfer the hierarchical features representation into a feature vector. Then a fully connected layer is added to reduce the dimensionality of the final output feature vector.

5.2. Long Term Temporal Modeling

The temporal modeling layer's goal in the proposed architecture is to predict PM_{10} concentration for the next 28 days in December 2015. The dataset for this particular duration is used as a test set and validates the model by comparing the prediction results of 28 days of December with the test dataset. In the temporal modeling layer, the input is the feature vector extracted from the CNN layer. The output of the CNN layer is treated as an input for Long Short Term Memory (LSTM) network to conduct temporal modeling of air pollution data. LSTM network is a long short term memory network that is suitable for dealing with the longer sequences time series dataset. LSTM differs from RNN due to the addition of a processor, which is utilized to judge the usefulness of the information. The structure of this processor is known as a cell. Usually, LSTM contains three gates in a cell, i.e., input gate, forget gate, and output gate.

The forget controls selectively how much information need to forget from the current cell. The output gate computes the output information, i.e., predicted PM_{10} concentration level. Input, output and forget can be implemented by using the following formulas,

$$i_t = \sigma(u_i h_{(t-1)} + w_i x_t + b_i) \tag{5}$$

$$c_t = f_t \odot c_{(t-1)} + i_t \odot tanh(u_c h_{(t-1)}) + w_c x_t + b_c) \tag{6}$$

$$f_t = \sigma(u_f h_{(t-1)} + w_f x_t + b_f) \tag{7}$$

$$o_t = \sigma(u_o h_{(t-1)} + w_o x_t + b_o) \tag{8}$$

$$h_t = o_t \odot tanh(c_t) \tag{9}$$

where σ is the element-wise activation function. i_t, f_t, o_t are the input gate, forget gate and output gate respectively. c_t, h_t represents cell state and hidden state vectors. u_i, u_c and u_o represents weight metrics for hidden state h_t, whereas w_i and w_c, w_o and w_f are the weight matrices of other gates. b_i, b_f, b_o are the bias vector for input gate, forget gate and output gate respectively. The basic structure of LSTM is presented in Figure 2.

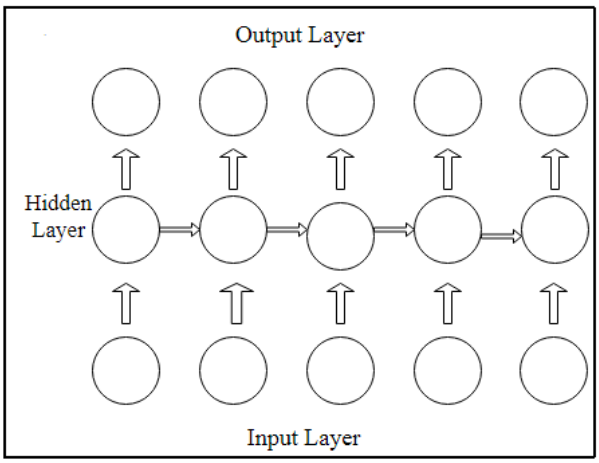

Figure 2. LSTM Structure

This paper utilized the Bidirectional LSTM (BILSTM) model (Verma et al. 2018; Graves and Schmidhuber

2005; Sun et al. 2019), which exhibits bidirectional properties of the LSTM network, where both past and future data play an important role in time series analysis. Two LSTM units stacked over each other in the BILSTM model (forward and backward). This model is capable enough to handle long-term dependencies without knowing prior information about past and future data. The stacked LSTM layer helps to capture hierarchical features in the temporal domain very efficiently. The output of each BILSTM layer fed into a fully connected layer. This is usually a dense presentation. This process continued for each time step. The output of the final timestamp will give the time series prediction results for each monitoring site. So implementing this model could provide better time series prediction accuracy. The basic structure of the BILSTM unit is shown in Figure 3.

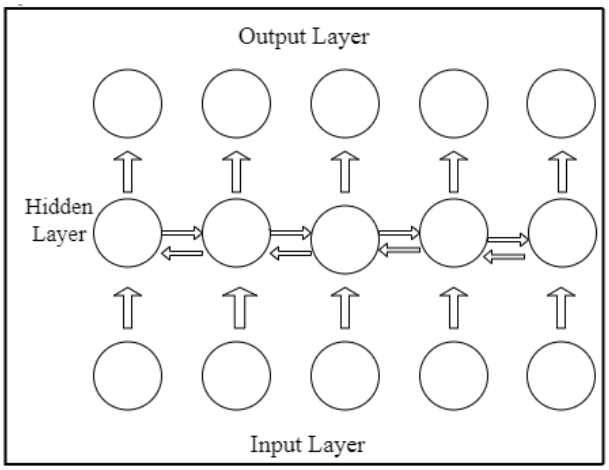

Figure 3. BILSTM Structure

It propagates the data through both directions i.e., called forward propagation and backward propagation. The forward propagation of time series data in BILSTM layer where time t ranges from 1 to T, can be formulated as below,

$$\overrightarrow{i_t} = \sigma(\overrightarrow{u_i}\,\overrightarrow{h_{(t-1)}} + \overrightarrow{w_i}\,\overrightarrow{x}_t + \overrightarrow{b_i}) \tag{10}$$

$$\overrightarrow{c_t} = \overrightarrow{f_t} \odot \overrightarrow{c_{(t-1)}} + \overrightarrow{i_t} \odot tanh(\overrightarrow{u_c}\,\overrightarrow{h_{(t-1)}}) + \overrightarrow{w}_c\,\overrightarrow{x_t} + \overrightarrow{b_c}) \tag{11}$$

$$\overrightarrow{f_t} = \sigma(\overrightarrow{u_f}\,\overrightarrow{h_{(t-1)}} + \overrightarrow{w_f}\,\overrightarrow{x_t} + \overrightarrow{b_f}) \tag{12}$$

$$\overrightarrow{o_t} = \sigma(\overrightarrow{u_o}\,\overrightarrow{h_{(t-1)}} + \overrightarrow{w_o}\,\overrightarrow{x_t} + \overrightarrow{b_o}) \tag{13}$$

$$\overrightarrow{h_t} = \overrightarrow{o_t} \odot tanh(\overrightarrow{c_t}) \tag{14}$$

The left arrow denotes the forward process. During backward propagation of time series data in BILSTM layer time, t ranges from T to 1. The backward propagation operations, represented by the right arrow, can be formulated using Equation 15-19.

$$\overleftarrow{i_t} = \sigma(\overleftarrow{u_i}\,\overleftarrow{h_{(t-1)}} + \overleftarrow{w}_i\,\overleftarrow{x_t} + \overleftarrow{b_i}) \tag{15}$$

$$\overleftarrow{c_t} = \overleftarrow{f_t} \odot \overleftarrow{c_{(t-1)}} + \overleftarrow{i_t} \odot tanh(\overleftarrow{u}_c\,\overleftarrow{h_{(t-1)}}) + \overleftarrow{w_c}\,\overleftarrow{x}_t + \overleftarrow{b_c}) \tag{16}$$

$$\overleftarrow{f}_t = \sigma(\overleftarrow{u_f}\,\overleftarrow{h_{(t-1)}} + \overleftarrow{w_f}\,\overleftarrow{x}_t + \overleftarrow{b_f}) \tag{17}$$

$$\overleftarrow{o_t} = \sigma(\overleftarrow{u_o}\,\overleftarrow{h_{(t-1)}} + \overleftarrow{w_o}\,\overleftarrow{x}_t + \overleftarrow{b_o}) \tag{18}$$

$$\overleftarrow{h_t} = \overleftarrow{o_t} \odot tanh(\overleftarrow{c}_t) \tag{19}$$

The final hidden element h_t can computed by using Equation 20

$$h_t = \overrightarrow{h_t} \odot \overleftarrow{h_t} \tag{20}$$

where $\overrightarrow{h_t}$, $\overleftarrow{h_t}$ denote forward out and backward output respectively.

5.3. spatial modeling

Monitoring air quality concentration level at each corner of a location and implement those data for further analysis to determine the air pollution impact on public health in a particular area are the essential steps in government initiated smart cities. The Indian government could establish a few air quality monitoring stations due to the high construction cost of monitoring sites and low government budget. Therefore, it is challenging to predict the air quality level at each location of a study area. It arises the necessity of spatial modeling of air pollution for air quality prediction. It helps to trace the harmful effect of pollution over a particular location. Identifying the spatial distribution of pollutant concentration for the present time may not be useful all the time; instead, it will be helpful if it can be predicted for the future. Hence, the IDW layer is added to the top of the CNN-BILSTM temporal modeling layer. CNN-BILSTM layer predicts air pollution levels for the existing 16 monitoring sites for the next four weeks of December 2015. Then the IDW layer of the model could predict the spatial distribution of pollutants by utilizing that predicted output of the CNN-BILSTM layer.

This paper used a distance-based spatial interpolation, IDW (Bhunia, Shit, and Maiti 2018; Gorai, Tchounwou, and Mitra 2017; X. Xie et al. 2017; Contreras and Ferri 2016) layer on the top of the temporal prediction layer to get the prediction value of each spatial distribution. It can be formulated as follows:

$$\hat{Z}(s_{t_0}) = \frac{\sum_{i=0}^{N} w_i s_{t_i}}{\sum_{i=0}^{N} w_i} \tag{21}$$

where, $\hat{Z}(s_{t_0})$ is the predicted value for unsampled point s_{t_0}, $w_i s_{t_i}$ is the weight function and and s_{t_i} is the sampled point value. In case of IDW, weight is the function of distance and can be estimated as follows:

$$w_i = \frac{1}{d(s_{t_0}, s_{t_i})^p} \tag{22}$$

Where the distance between measured point s_{t_i} and unmeasured point s_{t_0} denoted as d and the power factor as p. As closer points often have similar characteristics in space as per the first law of geography, those neighborhood points have more influence on interpolated point s_{t_0} values. The power value signifies the impact of neighborhood points on the interpolated point. The higher the power values, the more is the influence of neighborhood points on unsampled ones.

The pseudocode of the CNN-BILSTM-IDW algorithm is presented in Algorithm 1. The first part of the algorithm is meant to perform time series modeling followed by linear interpolation to handle missing values of the dataset. The second part of the algorithm conducts spatial modeling using IDW interpolation.

Algorithm 1 Spatio-temporal prediction algorithm for PM_{10}

Input: Air quality pollutant time series dataset $PM = [PM_1, PM_2.....PM_T]$, latitude and longitude (x,y), data sampling time t

Output: Pollutant prediction map

Initialization: Training process of CNN-BILSTM model with parameters ϕ.

for $1 \leq x, y \leq N$ **do**
  for $t = 1$ to T **do**
    if PM has missing value for duration t **then**
      Conduct linear interpolation
    else
      $PM_{t+d} \leftarrow (x, y)$ prediction of PM_{10} level for N number of monitoring sites by CNN-BILSTM model.
      return PM_{t+d}
    end if
  end for
end for
Generate spatio-temporal prediction map of PM_{10} for study area:
if (m, n) are the latitude and longitude of non-monitoring sites **then**
  return $\hat{P}M_{t+d}{}^{m,n} \leftarrow \sum_{x,y=1}^{N} w(x, y) * PM_{t+d}{}^{x,y}$
else
  return $PM_{t+d}{}^{x,y}$
end if

6. Results and Discussions

To evaluate the performance of the CNN-BILSTM-IDW model, we compared the performance of this model with ordinary IDW (Ya'acob et al. 2016), Exponential Kriging (EK) (Son, Bell, and J.-T. Lee 2010), Gaussian Kriging (GK), Spherical Kriging (SK) (Gong, Mattevada, and O'Bryant 2014), Universal Kriging (UK) (Son, Bell, and J.-T. Lee 2010), Radial Basis Function (RBF) (Bhunia, Shit, and Maiti 2018) and Empirical Bayesian Kriging (EBK) (Krivoruchko and Gribov 2019) models as shown in Table 1. Root Mean Square Error (RMSE) and Mean Error (ME) indicators are utilized to evaluate the performance of the CNN-BILSTM-IDW model. RMSE and ME can be calculated using Equation 23-24,

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left[\hat{Z}(x_i) - Z(x_i)\right]^2} \tag{23}$$

$$\text{ME} = \frac{1}{N}\sum_{i=1}^{n}\left[\hat{Z}(x_i) - Z(x_i)\right] \tag{24}$$

where, $\hat{Z}(xi)$ is the predicted value at location (x_i), $Z(x_i)$ is the observed value at (x_i) and N are the total number of monitoring stations. RMSE value of IDW, EK, RBF, GK, EBK, SK, UK, CNN-BILSTM-IDW model reduced (25.94), (24.64), (24.62), (24.48), (22.44), (24.38), (24.37), (21.71) respectively. Table 1 shows that the proposed spatio-temporal prediction model has better prediction performance than the other spatial prediction models.

Table 1. Model cross validation

Method type	Model	RMSE	ME
Deterministic	IDW	25.94	4.77
Geostatistics	EK	24.64	3.73
Geostatistics	GK	24.48	3.54
Geostatistics	SK	24.38	3.52
Geostatistics	UK	24.37	3.52
Geostatistics	EBK	22.44	3.53
Machine learning	RBF	24.62	3.80
Deep learning	CNN-BILSTM-IDW	21.71	3.50

Prediction maps generated by CNN-BILSTM-IDW for the four weeks of December 2015 are represented in Figure 4-7. Color scale indicates the variation of air pollution levels over the study area.

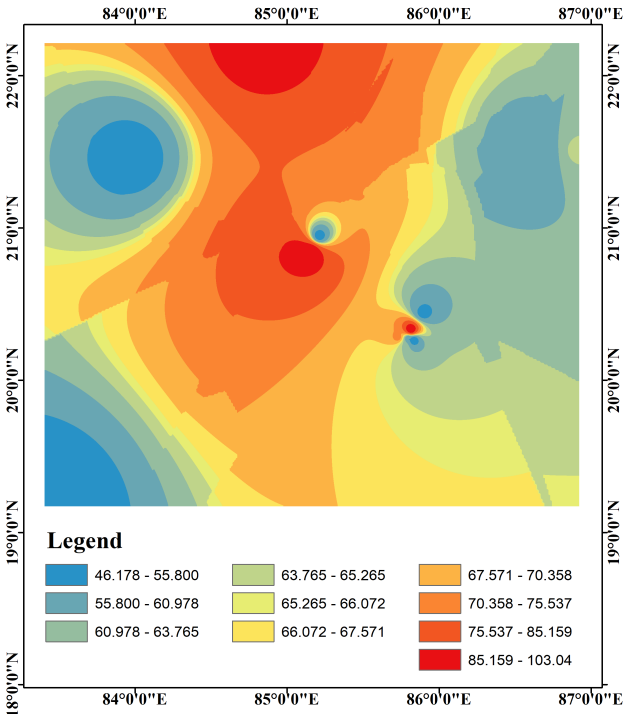

Figure 4. First-week prediction map. Spatial distribution of average PM_{10} value in the first week of December 2015.

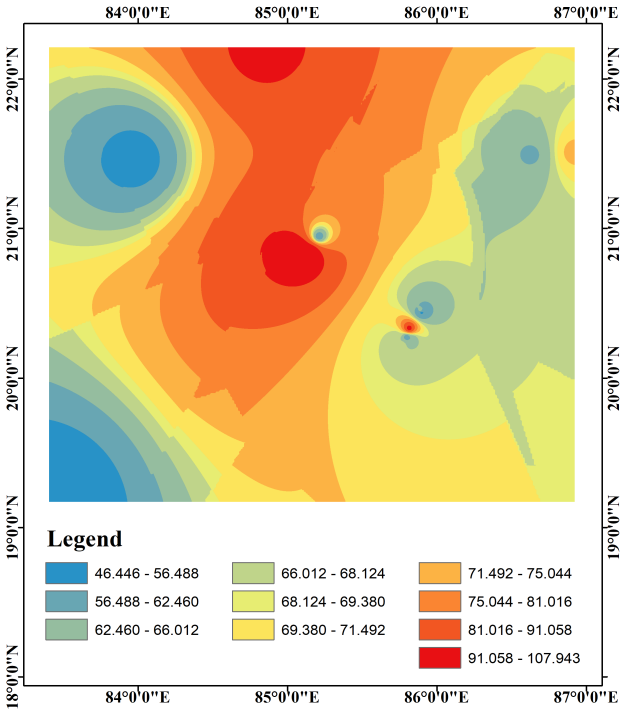

Figure 5. Second-week prediction map. Spatial distribution of average PM_{10} value in the second week of December 2015.

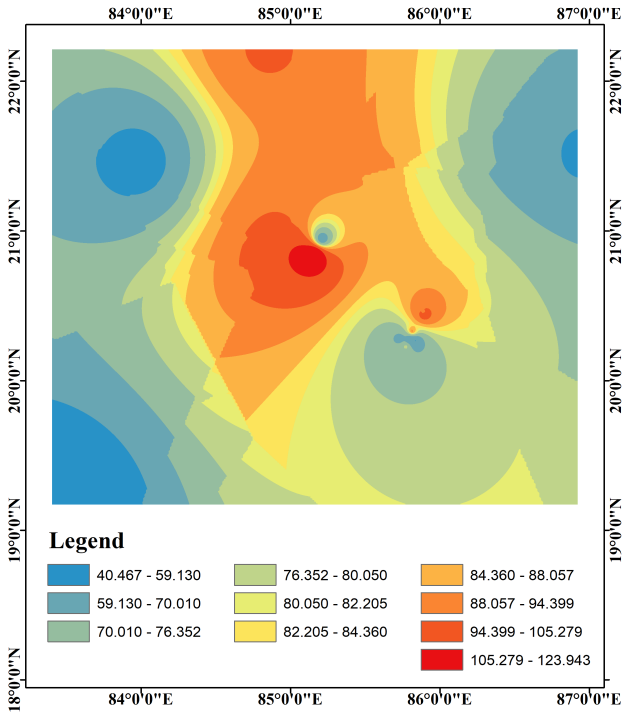

Figure 6. Third-week prediction map. Spatial distribution of average PM_{10} value in the third week of December 2015.

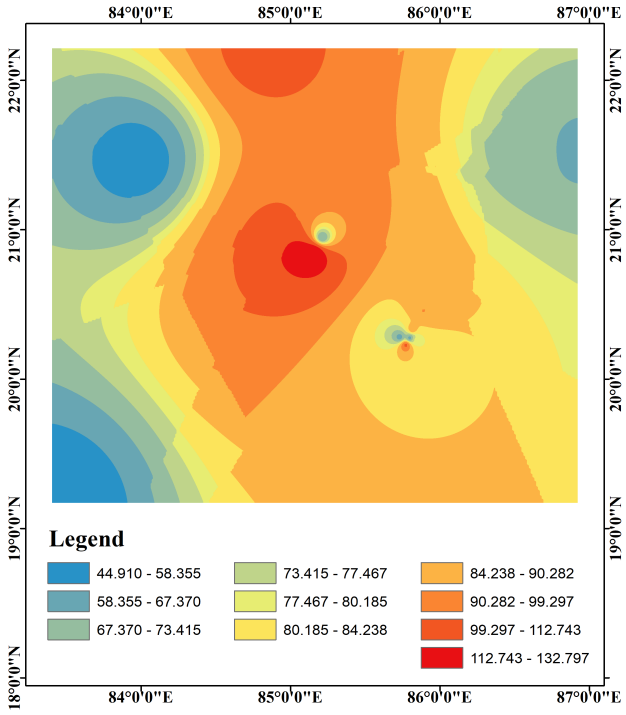

Figure 7. Fourth-week prediction map. Spatial distribution of average PM_{10} value in the fourth week of December 2015.

Few conclusions are derived from the above prediction maps of Odisha:
As presented in Figure 4-7, the eastern part of Odisha is predicted as a highly polluted area during December 2015, where PM_{10} concentration ranges from 67-132(µg/m3). That might be due to improper human activity, biomass burning, coal fields, and road transport emissions. It can be seen that the last week of December has the highest concentration level than the previous weeks. The results prove that the CNN-BILSTM-IDW model predicts air pollution levels not only for the current time but also for the future and the entire area by solving data imputation issues. It can be essential information for smart city users to take necessary preventive steps.

Figure 8-10 presents the user-friendly designed mobile application and web application to show the spatial prediction map of PM_{10} at a different period in advance. The spatial prediction maps are generated using the proposed CNN-BILSTM-IDW model. These user-end applications can be accessed from anywhere to get alert about the air quality level. These user applications are developed by Web App Builder of ArcGIS software, which provides location-based service accessibility. These services can also be used to access location information of treatment facilities and emergency services (Mbuh et al. 2020). The geo-enabled, IoT based dynamic end-user applications facilitate the decision-making process by improving situational awareness.

7. Conclusion

To conclude, this research paper proposed a new methodology framework that combines both deep learning and geostatistical approach to improve spatial prediction accuracy at a larger temporal granularity. The neural network layer improved the temporal prediction accuracy, whereas the IDW interpolation layer improved the spatial prediction accuracy in the study area.

This research work is conducted using only PM_{10} pollutant data due to proper data unavailability. Analyzing the influence of meteorological and traffic parameters on the ambient air quality could further improve the model prediction performance. In the future, if more data will be available, then using multivariate interpolation technique is expected to improve the prediction results.

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.

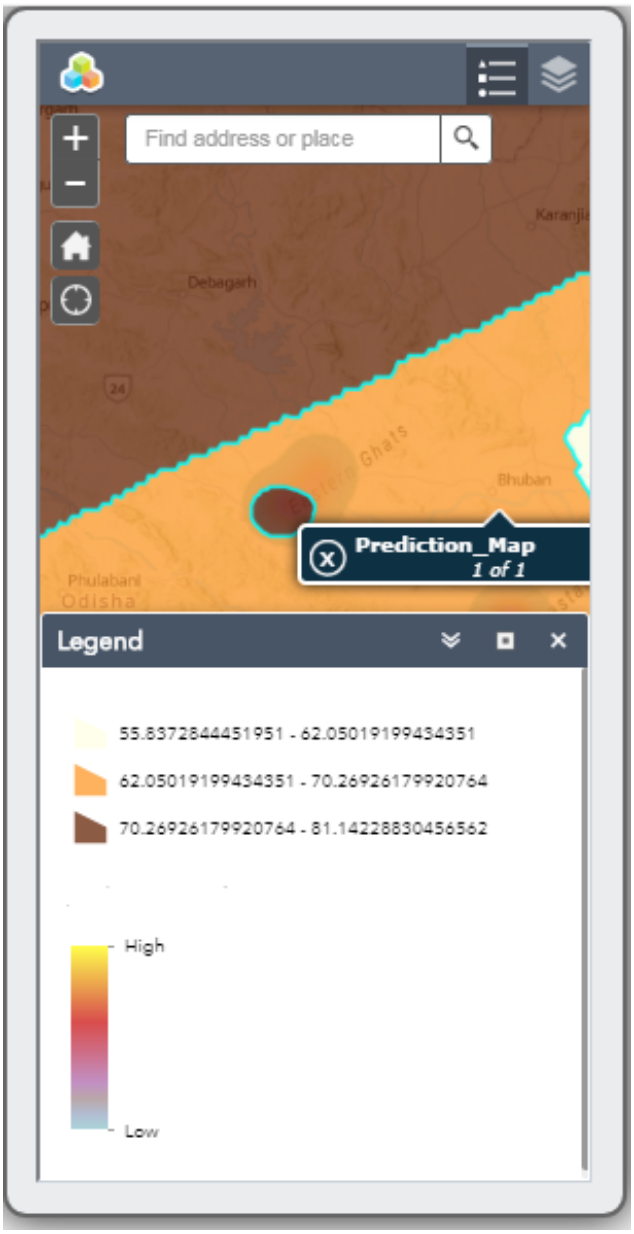

Figure 8. Developed a mobile application to present the air pollution spatial prediction map.

References

Beig, Gufran et al. (2020). "Objective evaluation of stubble emission of North India and quantifying its impact on air quality of Delhi". In: *Science of The Total Environment* 709, p. 136126.

Bhunia, Gouri Sankar, Pravat Kumar Shit, and Ramkrishna Maiti (2018). "Comparison of GIS-based interpolation methods for spatial distribution of soil organic carbon (SOC)". In: *Journal of the Saudi Society of Agricultural Sciences* 17.2, pp. 114–126.

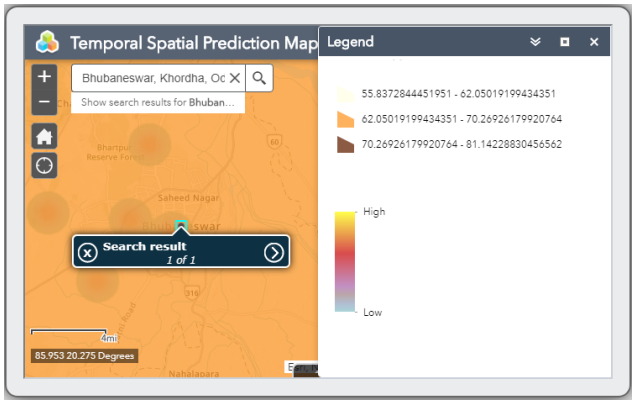

Figure 9. Developed a tablet application to present the air pollution spatial prediction map.

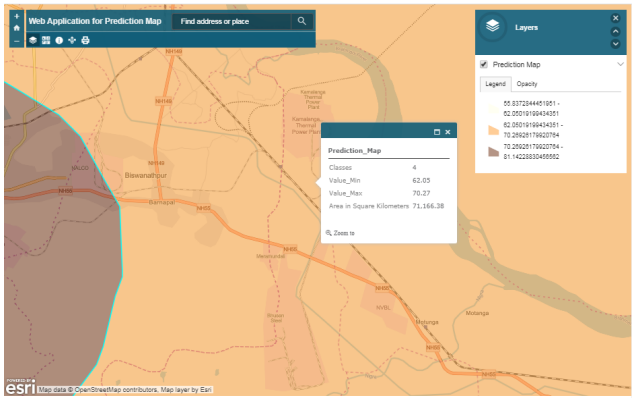

Figure 10. Developed a web client application to present the air pollution spatial prediction map.

Boys, BL et al. (2014). "Fifteen-year global time series of satellite-derived fine particulate matter". In: *Environmental science & technology* 48.19, pp. 11109–11118.

Brauer, Michael et al. (2019). "Examination of monitoring approaches for ambient air pollution: A case study for India". In: *Atmospheric Environment* 216, p. 116940.

Chowdhury, Sourangsu et al. (2019). "Tracking ambient PM2. 5 build-up in Delhi national capital region during the dry season over 15 years using a high-resolution (1 km) satellite aerosol dataset". In: *Atmospheric Environment* 204, pp. 142–150.

Contreras, Lidia and Cesar Ferri (2016). "Wind-sensitive interpolation of urban air pollution forecasts". In: *Procedia Computer Science* 80, pp. 313–323.

Contreras-Ochando, Lidia and Cesar Ferri (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, pp. 1296–1299.

Cressie, Noel (1990). "The origins of kriging". In: *Mathematical geology* 22.3, pp. 239–252.

Du, Shengdong et al. (2019). "Deep Air Quality Forecasting Using Hybrid Deep Learning Framework". In: *IEEE Transactions on Knowledge and Data Engineering*.

Elangasinghe, MA et al. (2014). "Complex time series analysis of PM10 and PM2. 5 for a coastal site using artificial neural network modelling and k-means clustering". In: *Atmospheric Environment* 94, pp. 106–116.

Fan, Junxiang et al. (2017). "A spatiotemporal prediction framework for air pollution based on deep rnn". In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 4, p. 15.

Feng, Xiao et al. (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, pp. 3029–3032.

Fu, Rui, Zuo Zhang, and Li Li (2016). "Using LSTM and GRU neural network methods for traffic flow prediction". In: *2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC)*. IEEE, pp. 324–328.

Gong, Gordon, Sravan Mattevada, and Sid E O'Bryant (2014). "Comparison of the accuracy of kriging and IDW interpolations in estimating groundwater arsenic concentrations in Texas". In: *Environmental research* 130, pp. 59–69.

Gorai, Amit Kumar, Paul B Tchounwou, and Gargi Mitra (2017). "Spatial variation of ground level ozone concentrations and its health impacts in an urban area in India". In: *Aerosol and air quality research* 17.4, p. 951.

Graves, Alex and Jürgen Schmidhuber (2005). "Framewise phoneme classification with bidirectional LSTM and other neural network architectures". In: *Neural networks* 18.5-6, pp. 602–610.

Gulliver, John et al. (2011). "Comparative assessment of GIS-based methods and metrics for estimating long-term exposures to air pollution". In: *Atmospheric environment* 45.39, pp. 7072–7080.

Gunarathna, MHJP, MKN Kumari, and KGS Nirmanee (2016). "Evaluation of interpolation methods for mapping pH of groundwater". In: *International journal of latest technology in engineering, management & applied science* 3, pp. 1–5.

Hama, Sarkawt ML et al. (2020). "Four-year assessment of ambient particulate matter and trace gases in the Delhi-NCR region of India". In: *Sustainable Cities and Society* 54, p. 102003.

Huang, Chiou-Jye and Ping-Huan Kuo (2018). "A deep cnn-lstm model for particulate matter (PM2. 5) forecasting in smart cities". In: *Sensors* 18.7, p. 2220.

Krivoruchko, Konstantin and Alexander Gribov (2019). "Evaluation of empirical Bayesian kriging". In: *Spatial Statistics* 32, p. 100368.

Lee, Muhammad Hisyam et al. (2012). "Seasonal ARIMA for forecasting air pollution index: A case study". In: *American Journal of Applied Sciences* 9.4, pp. 570–578.

Lee, Nam-Uk et al. (2018). "Design and implementation of the SARIMA–SVM time series analysis algorithm for the improvement of atmospheric environment forecast accuracy". In: *Soft Computing* 22.13, pp. 4275–4281.

Mbuh, Mbongowo et al. (2020). "Application of real-time GIS analytics to support spatial intelligent decision-making

in the era of big data for smart cities". In: *EAI Endorsed Transactions on Smart Cities* 4.9.

McKendry, Ian G (2002). "Evaluation of artificial neural networks for fine particulate pollution (PM10 and PM2. 5) forecasting". In: *Journal of the Air & Waste Management Association* 52.9, pp. 1096–1101.

Nayak, Tapaswini and Indrani Roy Chowdhury (2018). "Health damages from air pollution: Evidence from open cast coal mining region of Odisha, India". In: *Ecology* 1.1, pp. 42–66.

Nevtipilova, Veronika et al. (2014). "Testing artificial neural network (ANN) for spatial interpolation". In: *Journal of Geology & Geophysics* 3.2, pp. 01–09.

O'Shea, Keiron and Ryan Nash (2015). "An introduction to convolutional neural networks". In: *arXiv preprint arXiv:1511.08458*.

Odisha Information Technology Department, Odisha (2017, Oct 16). *Ambient Air Quality Data of Odisha*. URL: https://data.gov.in/.

Pan, Bingyue (2018). "Application of XGBoost algorithm in hourly PM2. 5 concentration prediction". In: *IOP Conference Series: Earth and Environmental Science*. Vol. 113. 1. IOP Publishing, p. 012127.

Pant, Pallavi, Sarath K Guttikunda, and Richard E Peltier (2016). "Exposure to particulate matter in India: A synthesis of findings and future directions". In: *Environmental research* 147, pp. 480–496.

Safavian, S Rasoul and David Landgrebe (1991). "A survey of decision tree classifier methodology". In: *IEEE transactions on systems, man, and cybernetics* 21.3, pp. 660–674.

Saide, Pablo E et al. (2011). "Forecasting urban PM10 and PM2. 5 pollution episodes in very stable nocturnal conditions and complex terrain using WRF–Chem CO tracer model". In: *Atmospheric Environment* 45.16, pp. 2769–2780.

Samal, K Krishna Rani, Korra Sathya Babu, and Santos Kumar Das (2020). "ORS: The Optimal Routing Solution for Smart City Users". In: *Electronic Systems and Intelligent Computing*. Springer, pp. 177–186.

Samal, K Krishna Rani, Korra Sathya Babu, Santosh Kumar Das, et al. (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*, pp. 80–85.

Shaban, Khaled Bashir, Abdullah Kadri, and Eman Rezk (2016). "Urban air pollution monitoring system with forecasting models". In: *IEEE Sensors Journal* 16.8, pp. 2598–2606.

Sharma, Anu Rani, Shailesh Kumar Kharol, and KVS Badarinath (2010). "Influence of vehicular traffic on urban air quality–A case study of Hyderabad, India". In: *Transportation Research Part D: Transport and Environment* 15.3, pp. 154–159.

Son, Ji-Young, Michelle L Bell, and Jong-Tae Lee (2010). "Individual exposure to air pollution and lung function in Korea: spatial analysis using multiple exposure approaches". In: *Environmental research* 110.8, pp. 739–749.

Sun, Jinlong et al. (2019). "Behavioral modeling and linearization of wideband RF power amplifiers using BiLSTM networks for 5G wireless systems". In: *IEEE Transactions on Vehicular Technology* 68.11, pp. 10348–10356.

Suykens, Johan AK and Joos Vandewalle (1999). "Least squares support vector machine classifiers". In: *Neural processing letters* 9.3, pp. 293–300.

Tao, Qing et al. (2019). "Air Pollution Forecasting Using a Deep Learning Model Based on 1D Convnets and Bidirectional GRU". In: *IEEE Access* 7, pp. 76690–76698.

Tripathy, DP and TR Dash (2018). "Assessment of Particulate and Trace Element Pollution in Airborne Dust around a Highly Mechanized Opencast Coal Mine in Talcher, Odisha". In: *Journal of Mining Science* 54.4, pp. 697–708.

Verma, Ishan et al. (2018). "Air pollutant severity prediction using Bi-directional LSTM Network". In: *2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*. IEEE, pp. 651–654.

Vorapracha, Phatarapon et al. (2015). "A comparison of Spatial Interpolation Methods for predicting concentrations of Particle Pollution (PM10)". In: *International Journal of Chemical, Environmental and Biological Sciences* 3.4, pp. 302–306.

Voynikova, DS et al. (2015). "Studying the effect of meteorological factors on the SO2 and PM10 pollution levels with refined versions of the SARIMA model". In: *AIP Conference Proceedings*. Vol. 1684. 1. AIP Publishing LLC, p. 100005.

Wang, Baowei, Weiwen Kong, and Hui Guan (2019). "Air quality forcasting based on gated recurrent long short-term memory model". In: *Proceedings of the ACM Turing Celebration Conference-China*. ACM, p. 128.

Wang, Jianzhou, Tong Niu, and Rui Wang (2017). "Research and application of an air quality early warning system based on a modified least squares support vector machine and a cloud model". In: *International journal of environmental research and public health* 14.3, p. 249.

Wang, Z et al. (2001). "A nested air quality prediction modeling system for urban and regional scales: Application for high-ozone episode in Taiwan". In: *Water, Air, and Soil Pollution* 130.1-4, pp. 391–396.

Xie, Haofei et al. (2019). "Research of PM2. 5 Prediction System Based on CNNs-GRU in Wuxi Urban Area". In: *IOP Conference Series: Earth and Environmental Science*. Vol. 300. 3. IOP Publishing, p. 032073.

Xie, Xingzhe et al. (2017). "A review of urban air pollution monitoring and exposure assessment methods". In: *ISPRS International Journal of Geo-Information* 6.12, p. 389.

Ya'acob, Norsuzila et al. (2016). "Haze monitoring based on air pollution index (API) and geographic information system (GIS)". In: *2016 IEEE Conference on Systems, Process and Control (ICSPC)*. IEEE, pp. 7–11.

Yenidoğan, Işil et al. (2018). "Bitcoin forecasting using arima and prophet". In: *2018 3rd International Conference on Computer Science and Engineering (UBMK)*. IEEE, pp. 621–624.

Zahedi, Gholamreza et al. (2014). "Ozone pollution prediction around industrial areas using fuzzy neural network approach". In: *CLEAN–Soil, Air, Water* 42.7, pp. 871–879.

Zamani Joharestani, Mehdi et al. (2019). "PM2. 5 Prediction Based on Random Forest, XGBoost, and Deep Learning Using Multisource Remote Sensing Data". In: *Atmosphere* 10.7, p. 373.

Zhou, Shanshan, Wenjing Li, and Junfei Qiao (2017). "Prediction of PM2. 5 concentration based on recurrent fuzzy neural network". In: *2017 36th Chinese Control Conference (CCC)*. IEEE, pp. 3920–3924.

Zhu, Min et al. (2018). "Class weights random forest algorithm for processing class imbalanced medical data". In: *IEEE Access* 6, pp. 4641–4652.

Zou, Bin et al. (2015). "Spatial modeling of PM 2.5 concentrations with a multifactoral radial basis function neural network". In: *Environmental science and pollution research* 22.14, pp. 10395–10404.