Spatial-temporal prediction of air quality by deep learning and kriging interpolation approach

K. Krishna Rani Samal¹, Korra Sathya Babu ², Santos Kumar Das ³

¹Vellore Institute of Technology, Vellore, India
²Indian Institute of Information Technology, Design and Manufacturing, Kurnool, India
³National Institute of Technology, Rourkela, India

Abstract

Air quality level is closely associated with our day-to-day life due to its serious negative impact on human health. Air pollution monitoring is one of the major steps of air pollution control and prevention. However, limited air pollution monitoring sites make it difficult to measure each corner of a region’s pollution level. This research work proposes a methodology framework incorporating a deep learning network, namely CNN-BIGRU-ANN and geostatistical Ordinary Kriging Interpolation model, to address this research gap. The proposed CNN-BIGRU-ANN time series prediction model predicts the $PM_{10}$ pollutant level for existing monitoring sites. Each monitoring site’s predicted output is transferred as input to the geostatistical Ordinary Kriging interpolation layer to generate the entire region’s spatial-temporal interpolation prediction map. The experimental results show the effectiveness of the proposed method in regional control of air pollution.

Received on 09 May 2023; accepted on 06 August 2023; published on 07 August 2023

Keywords: Deep learning, Transfer learning, Ordinary kriging, P M10

Copyright © 2023 K. K. R. Samal et al., licensed to EAI. This is an open access article distributed under the terms of the CC BY-NC-SA 4.0, which permits copying, redistributing, remixing, transformation, and building upon the material in any medium so long as the original work is properly cited.

doi:10.4108/eetsis.3325

1. Introduction

With the rapid growth of urbanization, people suffer from the rapid growth of urbanization, people suffer from the high concentration of air pollution causes lung cancer, respiratory diseases, premature death, and many skin-related health problems [2]. It also has a high negative impact on day to day activities and productivity. As per the World Health Organization (WHO) report, around seven million people die due to high exposure to air pollution every year [3]. Therefore, air pollution prediction in advance can help policymakers, the government, and the public in the environmental decision-making process to control air pollution levels.

Particulate matter ($PM_{2.5}$, $PM_{10}$) [4], sulfur dioxide ($SO_2$) [5], nitrogen dioxide ($NO_2$) [6], ozone ($O_3$), and carbon dioxide ($CO_2$) [7] are categorized as the most critical pollutants. However, particulate matter is the most dangerous contaminants due to its serious negative impact on the environment and public health [8]. Particulate matter usually includes dust from the construction site, industrial sources, and waste material burning. It is easy to inhale, but breathing particulate matter can cause various health problems [9] for pregnant women, older adults, and children. So, policymakers need to predict particulate matter more accurately.

$PM_{10}$ (Particulate matter with a diameter of less than 10 micrometers) [10] is one of the most dangerous pollutants among all the contaminants [11]. This pollutant concentration has become a severe issue worldwide [12]. Therefore, the control of $PM_{10}$ can improve the air quality level and protect human health and reduce economic loss. So the accurate prediction of the $PM_{10}$ level for the entire area can help people avoid this type of serious issue and allow people and the government to take necessary preventive action in advance [13].

Many air pollution monitoring stations have established at several locations to monitor air quality levels. Researchers and policymakers use sensor-generated data to predict air pollution levels for the next hour, the next day [14]. People can plan their daily activities to prevent the harmful impact of air pollution using this information. However, the lack of pollution information at each point can cause many discrepancies in the prediction results, which arises due to limited monitoring
sensors. This is typically a spatial interpolation and extrapolation issue of air pollution prediction [15].

Air pollution prediction problems generally include knowledge based and physical-based approach [16]. The physical prediction approach is based on atmospheric science and requires a strong knowledge of environmental science and pollutant diffusion mechanism [17]. Due to the complex pollutant diffusion mechanism, this approach has limited usage.

The knowledge-based prediction models are based on the dataset’s hidden properties. The knowledge-based prediction models include statistical, machine learning, and deep learning-based approaches. Traditional statistical prediction models fail to analyze the effect of other parameters on air pollutants. However, in the past studies, it is proven that the particular matter is widely affected by meteorological [18], climate variables, and traffic emissions [19]. Many research studies have experimented with data-driven based machine learning techniques and found better prediction results than the statistical model. But these models cannot process a vast amount of nonlinear multidimensional data sets. It is also challenging to perform correlation analysis among climate variables and air pollutants using machine learning-based techniques. Recently deep neural network techniques have been applied to overcome this issue [20]. With the rapid development of artificial intelligence techniques, deep learning-based methods are widely using for air quality modeling. Deep learning models, having many hidden layers, have proven better models for complex feature representation. These models have shown better generalization properties than machine learning-based models. The neural network’s hierarchical properties can better at deeper hidden feature extraction of air quality input data, developing better air pollution prediction models. Deep learning techniques have become the fastest-growing research field due to their high efficiency of analyzing the extensive time-series dataset’s temporal pattern and predicting the air pollution level over different time resolution [14]. So, deep learning models can be appropriate for our research work.

A vital air quality modeling component is understanding the temporal and spatial variations in ambient air pollutant concentration; usually, the ambient air pollution level reaches its peak level in the industrial estate and congested traffic area [21]. It has been noticed from various research studies that small variations of ambient air pollution levels may occur over short distances. A monitoring site’s air pollution data can be considered representative of a small area within that distance. Still, the deep learning model only considers a particular monitoring site’s location during air quality prediction but does not capture the pollutant concentration variation due to local sources and other places [19].

On the other hand, geostatistical models are the traditional spatial prediction models that consider air pollution’s spatial variability during air quality modeling [22]. It is notable that most geostatistical methods, including Kriging, Inverse Distance Weighting interpolation technique, do not analyze the temporal variation of extensive time-series data. From the research study, it is found that most of the prediction models either predict the spatial variation or temporal variation of air pollution on a larger scale but not at the same time. This research aims to predict pollutant concentration at a high temporal resolution and spatial variation with improved accuracy.

Following the introduction, the rest is organized as follows. Section 2 represents the literature review. Section 3 formulates the problem; Section 4 presents the study area. Section 5 focuses on the proposed methodology framework. The results are discussed in Section 6, and Section 7 concludes the research study.

2. Literature Review

Air pollution prediction has been conducted using data-driven approaches like Autoregressive Integrated Moving Average (ARIMA) [23] and the Seasonal Autoregressive Integrated Moving Average (SARIMA) [24] model, which are based on data stationarity. Stationarity based existing prediction models are often violated by air pollution dataset. Later, nonlinear machine learning models such as Random Forest, Radial Basis Function (RBF) [25], Principle Component Analysis (PCA) [26], Support Vector Regression [25], Support Vector Machine (SVM) [27, 28], Artificial Neural Network (ANN) [29], BP Neural Network [30] came into the picture to perform time series prediction. On the other hand, with the rapid growth of AI technologies, nonlinear machine learning models are not used as state-of-the-art models because of the gradient decent issues. Deep learning models, e.g., Recurrent Neural Network (RNN) [31], Long Short Term Memory Network (LSTM) [32] and Gated Recurrent Units (GRU) models are reported as the better prediction models due to their efficiency of handling both short term and long term dependency in the high volume of time series dataset. These models can also overcome the gradient decent issues of nonlinear machine learning problems very efficiently [20].

FU et al. studied time series prediction using LSTM, GRU model and found that recurrent neural network-based model outperforms the univariate statistical time series prediction models [33]. Later on, the researcher studied to develop multivariate time series prediction model as these models are not suitable for multivariate time series prediction. Maggiolo et al. proposed a multivariate autoregressive convolutional
Spatial-temporal prediction of air quality by deep learning and kriging interpolation approach

recurrent neural network prediction model. This model is useful only in sequence to sequence modeling, not for nonlinear architecture [34]. Several other models also proposed for multivariate time series prediction like RNN-LSTM [35], Gated Recurrent Long Short Term Memory model [36] and Bidirectional LSTM (BILSTM) [37]. Few researchers applied bidirectional properties of the recurrent neural network to identify both forward and backward temporal dependency to improve the prediction performance. Bidirectional GRU (BIGRU) [38] model generates lower error value than GRU and hence, exhibits better accuracy during PM\textsubscript{2.5} time series prediction. Moreover, this model outperforms the other existing model; furthermore, it gives better results than the baseline LSTM model due to its bidirectional properties. Ge et al. developed the DBU-LSTM model based on the unidirectional and bidirectional properties of RNN [39]. The DBU-LSTM model was evaluated with a real-time air quality dataset of Beijing, which outperforms the unidirectional LSTM model. These time series-based prediction models are not efficient enough to analyze the dataset features’ hidden correlation, affecting the prediction results. So, Huang et al. adopted both CNN and RNN based LSTM architecture to get a good quality of PM\textsubscript{2.5} prediction value, where CNN identifies the correlation among the feature and reduce the dimensionality of the dataset and LSTM is used for temporal air quality modeling [40]. These models have proved better at long-term dependency modeling. These studies have some limitations as they do not consider the spatial variability of air pollution concentration.

Some research studies improved the prediction performance of time series prediction models by integrating spatial features. Spatial-temporal prediction problems depend upon the characteristics of feature value and its geographical coordinates. The spatial-temporal prediction problem for irregular grid [41] is the most challenging issue in the deep learning research area. The most challenging part is to model the temporal and spatial dependency accurately within the data. Sun et al. put forward Inverse Distance Weighting (IDW) interpolation and data diffusion method to make full use of time, space [42]. They proposed a spatial-temporal prediction framework to get accurate results. Extracting spatial characteristics in a high dimension is a significant issue as the interpolation techniques depend upon the assumption that the research study objective is static. To overcome these limitations, few CNN based architecture was adopted by a few researchers. Xie et al. analyzed spatial characteristics of PM\textsubscript{2.5} monitoring stations and developed CNN’s GRU model to extract spatial features of multi-scale data in a high dimension automatically to develop an advanced PM\textsubscript{2.5} prediction framework [43]. Lin et al. extend the DCRNN model concept to establish GC-DCRNN (Geo Context-based Diffusion Convolutional Recurrent Neural Network) model. The GC-DCRNN model implements neighborhood characteristics and automatically extracts the essential factors affecting air pollution to find the best prediction result [44]. These models exhibited excellent prediction performance, handling temporal and spatial variability. Most of these models focus on the hourly temporal resolution or one step ahead prediction. Air pollution prediction for the next day, next week, and the following month are more useful. However, a minimal study has conducted predicting air pollution levels at a high temporal and spatial resolution and has seldom been solved.

Based on the above survey, this paper proposes an algorithm based on recurrent neural networks and a spatial modeling approach to automatically extract the features and identify long-term dependency to get better spatial-temporal air pollution prediction results. It replaces the traditional neural network-based prediction models, which cannot simultaneously solve feature extraction, spatial-temporal correlation, and data interpolation issues.

3. Problem formulation
Due to unstructured air pollution monitoring stations, it is difficult to identify the spatial distribution of air pollution levels and predict its value for an unknown point. In order to predict the temporal-spatial distribution of air pollution levels for an entire region, a proper prediction model should be developed followed by spatio-temporal data analysis. The pollution levels for existing monitoring stations can be utilized here to predict unknown points.

Given a set of historical air pollution dataset \( P = [p_1, p_2, p_3, ..., p_l] \) of time length \( t \) for each \((i \times j)\) monitoring stations within an area \( A \) for all \( 1 \leq i, j \leq N \). Each pollution monitoring stations has observed time-series dataset \( P \) for each pollutant i.e.,

\[
P = \begin{pmatrix}
p_{1_{(1,1)}} & \cdots & p_{1_{(1,j)}} \\
\vdots & \ddots & \vdots \\
p_{l_{(1,1)}} & \cdots & p_{l_{(1,j)}}
\end{pmatrix}
\]

where, \( i \) and \( j \) are the longitude and latitude of each pollution monitoring stations and treated as spatial features of dataset. Given \( P \) as the set of time series pollution dataset with time window \( t \) in a particular region, the objective is to predict the future pollution level \( p_{i+d}^{m,n} \) at unmeasured point \((m,n)\) for time \((t+d)\), where \( d \) is the number of next days for which it is necessary to know the future pollution level.

4. Study area
Odisha, an Indian state, is selected as the research area, which covers 4.87 percentage of the country...
and approximately 155,707 km² area. As ambient air pollution has become one of Odisha’s critical concerns [45], this state is considered the experiment area to evaluate the proposed model’s performance. The historical air pollution dataset [46] collected from an Indian government open-source website [47] used to model air pollution prediction algorithms. The distribution of air pollution monitoring stations in Odisha is represented in (Fig. 1). The dataset includes PM₁₀ pollution concentration value for 2004-2015 with station id, sampling locations, and sampling date as attributes. The dataset description is presented in Table 1.

Table 1. Dataset description

<table>
<thead>
<tr>
<th>Dataset input variables</th>
<th>Variable value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of monitoring stations</td>
<td>16</td>
</tr>
<tr>
<td>Time period</td>
<td>2004-2015</td>
</tr>
<tr>
<td>Time interval (In day)</td>
<td>1</td>
</tr>
<tr>
<td>Pollutant</td>
<td>PM₁₀</td>
</tr>
<tr>
<td>Pollutant Unit</td>
<td>µg/m³</td>
</tr>
<tr>
<td>Number of records</td>
<td>12616</td>
</tr>
<tr>
<td>Minimum value</td>
<td>56</td>
</tr>
<tr>
<td>Maximum value</td>
<td>202</td>
</tr>
<tr>
<td>Mean</td>
<td>125.57</td>
</tr>
<tr>
<td>Median</td>
<td>124.5</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>42.158</td>
</tr>
</tbody>
</table>

5. Proposed approach

The proposed model consists of four parts. One dimensional CNN (1D convNet) is used for feature extraction and reduces dimensionality during the initial stage. In the second step, extracted features fed into the bi-directional GRU model for time series prediction for all sampled points. Though the CNN-BIGRU model gives good quality prediction results, there is an opportunity to maximize its accuracy. It is possible to analyze the error component as a nonlinear component and model it with a nonlinear artificial neural network (ANN) [48] to improve accuracy further. At the end stage, the time series model’s output neurons are fed into the geostatistical Spherical Ordinary Kriging (OK) [22] layer to improve its spatial prediction capability. Thus, the proposed method performs spatial-temporal prediction in the study area. The proposed architecture is represented in (Fig. 2). Each part of the proposed architecture is discussed in the below subsections.

5.1. Preprocessing steps

The collected ground observations sensor dataset contains PM₁₀ concentration level of 16 air pollution monitoring sites. The day-wise sampled data are normalized, and outliers are removed in the preprocessing step to improve the performance and increase the model’s learning speed. The z-score normalization technique [49] is utilized to normalize air pollution dataset. As the collected dataset has missing attributes, this study adopted the temporal interpolation method to recover the missing attributes in the preprocessing step.

Feature extraction layer. After the preprocessing step, the normalized output is used as the input to the one-dimensional Convolutional Neural Network (1D CNN) layer. This layer is based on data sparsity and weight sharing function [46]. Though it has tremendous application in image classification, it is also useful for time series prediction due to its sequential modeling capabilities. This layer converts the input into a form that can be directly used in further stages. In 1D CNN, each hidden layer works as a convolution layer with input and weight as a vector parameter and can extract high-level and low-level features by reducing the data’s dimensionality. Max-pooling operation is conducted, followed by the convolutional process to minimize the number of parameters and computation of the framework by reducing the time series pollution dataset’s length. As time acts as a spatial dimension in convNet [50], the max-pooling operation also helps to minimize spatial dimensionality.

Temporal modeling layer. As the sensor data are usually in time series, air quality modeling is possible by analyzing its time series dependency pattern. The recurrent neural network based deep learning model, GRU [51], is implemented to predict PM₁₀ concentration value. GRU model is more straightforward than LSTM [52] architecture due to its less number of gates, i.e., update gate and reset gate [53]. The basic structure of the GRU model is presented in (Fig. 3).

The Update gate (zₜ) captures the long-term dependencies, whereas the reset gate (rₜ) captures the short term dependencies. (zₜ) decides how much information should keep, and the reset gate signifies the amount of data need to forget. It’s fewer parameter utilization is simpler to train and time-consuming, and efficient enough to solve vanishing gradient issues. The basic GRU model can be estimated as follows [50]:

\[ rₜ = σ(w_r \ast [xₜ, h_{t-1}]) \]  \hspace{1cm} (1)

\[ zₜ = σ(w_z \ast [xₜ, h_{t-1}]) \]  \hspace{1cm} (2)

\[ hₜ = (1 - zₜ) \ast h_{t-1} + zₜ \ast h_t \]  \hspace{1cm} (3)

\[  \hat{h}_t = σ(w_h \ast [xₜ, rₜ \ast h_{t-1}]) \]  \hspace{1cm} (4)

\((rₜ)\) process pollution dataset as input \((xₜ)\) and hidden state \((h_{t-1})\) and then implement sigmoid \((σ)\) activation.
Figure 1. Air pollution monitoring stations. The marked red area indicates the Odisha state situated in the country of India. Green marked dotted symbols represent the spatial distribution of air pollution monitoring stations in Odisha.

Figure 2. The proposed model system architecture. The figure shows the feature extraction layer (CNN), temporal modeling layer (BIGRU), fully connected layer, fine-tuning layer, and interpolation layer of the proposed model, which are involved in performing spatial-temporal air pollution prediction.
dependencies at a different timestamp. The unit, which is used for capturing short term dependencies, will manage to have a frequently active reset gate and active update gate for long term dependencies. The candidate activation function \( \hat{h}_t \) is estimated with \( r_t \). The actual activation function \( h_t \) of the model at time \( t \), is a linear interpolation between \( \hat{h}_t \) and \( h_{t-1} \). Furthermore, as both the past and future data play an essential role in temporal modeling, the experiment was conducted using the bidirectional GRU (BIGRU) model to conduct both forward and backward propagation [50]. This model can capture the temporal dependency from both directions and give more accurate prediction results due to its lower error, which is impossible in the one-directional GRU model. Further, with the combination of 1D convNet, it gives far better performance than any other simple recurrent neural network-based model. To further optimize the performance, fine-tuning is the ultimate solution, so we added a nonlinear recursive artificial neural network (ANN) [25] model, which adjusts the bias by finding the correlation between the outputs for better prediction results. This proved its effectiveness in temporal modeling of time-series datasets. The details of the experimental setting of temporal air quality modeling are represented in Table 2.

![Figure 3. Basis structure of GRU model [52].](image)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training set</td>
<td>80%</td>
</tr>
<tr>
<td>Testing set</td>
<td>10%</td>
</tr>
<tr>
<td>Validation set</td>
<td>10%</td>
</tr>
<tr>
<td>Epochs</td>
<td>2000</td>
</tr>
<tr>
<td>Optimizer</td>
<td>Adam</td>
</tr>
<tr>
<td>Learning rate</td>
<td>0.001</td>
</tr>
<tr>
<td>Activation function</td>
<td>Relu</td>
</tr>
<tr>
<td>Loss function</td>
<td>MSE</td>
</tr>
</tbody>
</table>

As discussed earlier, deep learning models are mostly used to perform time series forecasting for each site but unable to predict pollution levels for the entire region, including the places where there is no air pollution monitoring station [15]. This research paper proposes a spatial-temporal interpolation prediction approach, which utilizes the obtained time series prediction results of each site as its input and performs spatial prediction to overcome this research gap.

**Spatial–temporal interpolation using transfer learning approach.**

Air quality monitoring has a vital role in air pollution reduction, but minimal monitoring stations have established to monitor the air pollution level due to high construction costs. It is challenging to follow air quality levels at each corner of a location, which arises the need for spatial analysis of air quality to understand its harmful impact over a particular area [54]. It is also necessary to predict spatial distribution for the next few days to take appropriate preventive action against the severe adverse effects of pollution in advance.

The \( PM_{10} \) pollutant prediction of each site is done using the CNN-BIGRU-ANN model individually. Afterward, those predicted values were transferred as the input for Ordinary Kriging interpolation model to conduct both forward and backward propagation in the range of observation points to generate a continuous surface of \( PM_{10} \). Thus, this operation would help predict the \( PM_{10} \) level at each measured and unmeasured location to visualize the \( PM_{10} \) level for the entire study area.

Ordinary Kriging (OK) [55] model calculate the unknown point values using known point values with the help of a weight function. The weight function is the weighted average of the available data instead of just considering distance like the IDW [22] model. The weight calculation in kriging is based on the observations data’s assumptions, whether the process is second-order stationarity based or depends upon the covariates. The unmeasured point values can be predicted using the OK model as follows [56],
where $(m,n)$ is the point whose value needs to be predicted, $y(i,j)$ is the weight of the observed value at location $(i,j)$, $N$ is the number of sampling points in the neighborhood area and $z(i,j)$ is the measured value at location $(i,j)$. The OK model's weight function can be estimated using semivariance, which is usually used to determine the spatial dependence. Semivariance expresses the relationship between the existing data and the estimation point. It can be estimated as half the squared difference between the paired data point values. Semivariance value $\lambda$ at $h$ can be computed as [57],

$$\lambda(h) = \frac{1}{2nh} \sum_{i=1}^{n_k} [z(i,j) - z((i,j) + h)]^2$$ (6)

where, $h$ represents the distance between two points $(i,j)$ and $((i,j) + h)$. $n_k$ represents the number of sample points within the searching neighborhood area $h$ used to calculate the variance value between $(i,j)$ and $((i,j) + h)$. $z(i,j)$ represent the observed pollutant value at point $(i,j)$ and $z((i,j) + h)$ is the observed pollutant value at $(i,j) + h$ [58].

The OK model follows two steps to identify spatial distribution. The first one is to discover autocorrelation among spatial data, and the second one is for prediction. It considers adjacent points weighted by distance in the interpolated area and extent of autocorrelation $\lambda(h)$ to quantify optimum weights at each sampling distance [59]. After removing the data's spatial trend, it determines the best variogram model and then generates the required interpolated surface. The Ordinary Kriging model is utilized as the end layer of the proposed architecture to generate Odisha's prediction map over the next 28 days of December 2015.

6. Results and discussions

In this paper, the proposed architecture is designed for spatial-temporal prediction of $PM_{10}$ in the study area. This research paper considers the temporal correlation of air pollution data and examines the spatial relationship among the unmeasured locations to the monitoring sites during air quality modeling. The time series prediction result of the proposed CNN-BIGRU-ANN model for the 16 monitoring sites of Odisha is shown in (Fig. 4-5). (Fig. 4-5) shows the time series prediction of the input variable $PM_{10}$, where both observed and predicted values follow the same fluctuations for each site, which offers the effectiveness of the proposed CNN-BIGRU-ANN model in temporal modeling. The X-Axis presents the duration of the time series prediction, while the Y-Axis of the figures represents the predicted $PM_{10}$ value for that duration.

Another salient feature of the proposed method is that it has a newly designed Ordinary Kriging interpolation layer, which interpolates CNN-BIGRU-ANN time series prediction results to generate a spatial-temporal pollution prediction map. This is the most important aspect of air pollution management as it enables us to conduct air pollution prediction and map the correlated data successfully. The time series prediction results of these 16 air quality monitoring sites of Odisha are utilized to perform spatial interpolation using an Ordinary Kriging technique, which generates a spatial-temporal interpolation map for the entire study area. For the illustration, the spatial-temporal interpolation results of each day of December 2015 are presented in (Fig. 6-7). (Fig. 8) presents the designed web application to display the spatial prediction map of $PM_{10}$ at a high temporal resolution for the study area. Hence, the user can access the prediction information in advance and keep them safe by taking necessary preventive action.

From the interpolation prediction map, we can see a clear peak value of $PM_{10}$ in the eastern area of Odisha, i.e., the capital city area. That may be due to the high transportation network and inappropriate human activity. As per government statistical report, the total number of transportation and non-transportation registered vehicles has increased to 574564 and 5258793, respectively, as of 31st March 2016 [60] in Odisha. An increasing number of vehicle transportation might be the main reason for this high emission of $PM_{10}$ across all over Odisha, which is marked in red color in (Fig. 6-7). Still, further research requires to find out the reason behind the high emission of pollutants in the study area. Spatial prediction maps could be essential information for smart city users, policymakers, city planners, and the government to initiate the needful preventive service against the unfavorable situation of air pollution.

To further evaluate the performance, the proposed CNN-BIGRU-ANN model is compared with other baseline and state-of-art models. Statistical prediction model like SARIMA [46] and FbPROPHET [61] models are used as baseline models for performance comparison. Moreover, deep learning based LSTM, GRU, BIGRU, and bi-directional LSTM (BILSTM) models, are also used as state-of-arts models for comparison purposes. In the experiment, 80% of the data is used for training and 10% of data, i.e., last month data is used for testing purposes, and the model is trained to predict $PM_{10}$ value for the next 28 days of December 2015. Parameters of LSTM, GRU, BILSTM, and BIGRU models are optimized during the training process. Three metrics, including Root Mean
Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE) is utilized to evaluate the generated errors and determine the prediction quality of the proposed CNN-BIGRU-ANN prediction model. Used error metrics can be expressed as shown in Equation 7-9 where $i$ is the total error, $i=0,1,2...n$.

The smaller the RMSE, MAE, and MAPE value, the better the prediction result will be due to less prediction error.

\[
RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} e_i^2}
\]

\[
MAE = \frac{1}{n} \sum_{i=1}^{n} |e_i|
\]

\[
MAPE = \frac{100}{n} \sum_{i=1}^{n} \left| \frac{e(t)}{A(t)} \right|
\]

The comparative results of the models are represented in Table 3.

**Table 3.** Time series prediction performance comparison of CNN-BIGRU-ANN and other baseline models.

<table>
<thead>
<tr>
<th>Model</th>
<th>RMSE</th>
<th>MAE</th>
<th>MAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>PROPHET</td>
<td>61.984</td>
<td>46.667</td>
<td>68.90</td>
</tr>
<tr>
<td>BILSTM</td>
<td>56.260</td>
<td>47.738</td>
<td>80.016</td>
</tr>
<tr>
<td>SARIMA</td>
<td>51.925</td>
<td>20.985</td>
<td>24.560</td>
</tr>
<tr>
<td>BIGRU</td>
<td>45.969</td>
<td>34.056</td>
<td>58.261</td>
</tr>
<tr>
<td>LSTM</td>
<td>41.446</td>
<td>32.743</td>
<td>60.067</td>
</tr>
<tr>
<td>GRU</td>
<td>33.870</td>
<td>28.098</td>
<td>49.042</td>
</tr>
<tr>
<td>CNN-BIGRU-ANN</td>
<td>15.403</td>
<td>13.012</td>
<td>20.049</td>
</tr>
</tbody>
</table>

Table 3 shows that the FbPROPHET model has higher RMSE, MAE, and MAPE due to their nonlinear properties. BILSTM, SARIMA, BIGRU, LSTM, and GRU have relatively lower error metrics. The comparative analysis proved that the proposed CNN-BIGRU-ANN prediction model had the best performance, having the lowest error metric value. It exhibits the bidirectional
properties of RNN structure [4] and performs a fine-tuning operation to improve prediction results at a high temporal granularity.

It can easily understand that CNN-BIGRU-ANN is 75% better than the PROPHET in terms of RMSE value. Furthermore, it is also found that the CNN-BIGRU-ANN model is 62% better than ordinary LSTM and 54% better than the ordinary GRU model in air quality modeling over the 28 days. Table 3 proves that the proposed CNN-BIGRU-ANN is almost 54%-75% better than the other time series prediction models.

The CNN-BIGRU-ANN time series prediction results have experimented with different interpolation techniques like Radial Basis Function (RBF), Inverse Distance Weighting (IDW) [62], Simple Kriging (SK), Universal Kriging (UK), and Ordinary Kriging (OK) [56] for a fair comparison of model’s interpolation performance. Radial basis functions [63] are based on basic radial functions, which are generally used for creating the neural network. This interpolation is usually used to generate a surface using a large number of sampling points. IDW [64] is a non-statistics interpolation method and used to analyze the variation of the local surface. This method used the linear weighted function to find out each cell’s values of a local surface. The inverse distance ratio's power is used as a weight to calculate the unmeasured point’s value. Simple Kriging is one of the variants of Kriging techniques where local means are constant. Universal Kriging [65] is a variation of the Ordinary Kriging method with the non-stationary condition. In the UK model, the mean value differs at different locations in a deterministic way, while variance remains constant. Hence, the UK is a type of kriging method with a local trend or spatial drift.

Root Mean Square Error (RMSE) and Mean Error (ME) are used as error metrics to evaluate the interpolation performance of the proposed method, as
Figure 6. The temporal-spatial prediction map: Day-wise average $PM_{10}$ value at Odisha (1st December 2015-16 December 2015). The color indicates $PM_{10}$ prediction value over the layer, and the dotted symbols represent the air pollution monitoring stations of Odisha.

Figure 7. The temporal-spatial prediction map: Day-wise average $PM_{10}$ value at Odisha (17 December 2015-28 December 2015). The color indicates $PM_{10}$ prediction value over the layer, and the dotted symbols represent the air pollution monitoring stations of Odisha.
shown in Table 4. Table 4 shows that the Ordinary Kriging interpolates the CNN-BIGRU-ANN time series prediction results more efficiently than any other interpolation techniques due to lower RMSE and ME error metrics. Hence, the Ordinary Kriging technique is adopted to generate the Spatial-temporal prediction map for the study area.

<table>
<thead>
<tr>
<th>Model</th>
<th>RMSE</th>
<th>ME</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNN-BIGRU-ANN-RBF</td>
<td>23.26</td>
<td>5.47</td>
</tr>
<tr>
<td>CNN-BIGRU-ANN-IDW</td>
<td>23.25</td>
<td>4.73</td>
</tr>
<tr>
<td>CNN-BIGRU-ANN-SK</td>
<td>21.72</td>
<td>4.29</td>
</tr>
<tr>
<td>CNN-BIGRU-ANN-UK</td>
<td>22.65</td>
<td>4.51</td>
</tr>
<tr>
<td>CNN-BIGRU-ANN-OK</td>
<td>16.40</td>
<td>3.68</td>
</tr>
</tbody>
</table>

7. Conclusion and future work

People across developing countries like India face high exposure to air pollution due to various factors. Thus the accurate spatial-temporal prediction of pollutants is a crucial step to avoid high air pollution exposure risk. In this research paper, a deep learning-based CNN-BIGRU-ANN prediction model was first proposed to predict PM$_{10}$ level for existing monitoring sites. The proposed CNN-BIGRU-ANN has 75% better prediction results than the FbPROPHET model. The proposed time series prediction model uses the kriging layer to generate the entire study area's spatial-temporal prediction map. Experimental results show that the proposed method incorporated with deep learning and geostatistics approach could provide early information of pollutant level for a particular site and the entire region for reducing the health risk associated with PM$_{10}$ pollutant.

We can extend this research work in the future by adding more variables like meteorological factors and traffic, which significantly contribute to the rising air pollution level. We have utilized only PM$_{10}$ concentration level for spatial-temporal prediction due to data unavailability. Analyzing the effect of other factors on air pollution may further improve model performance.

8. Acknowledgment

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


[27] Ibbi, A., Kerchich, Y., Hadidi, N., Merabet, H. and Hentabli, M. (2020) Prediction of the concentrations of pm1, pm2.5, pm4, and pm10 by using the hybrid dragonfly-svm algorithm. Air Quality, Atmosphere & Health: 1–11.


