A Novel Temporal-spatial Interpolation Method for Spatio-temporal Air Quality Forecasting

1st K.Krishna Rani Samal
National Institute of Technology
Rourkela, India
517CS6019@nitrkl.ac.in

2nd Korra Sathya Babu
National Institute of Technology
Rourkela, India
ksathyababu@nitrkl.ac.in

3rd Santos Kumar Das
National Institute of Technology
Rourkela, India
dassk@nitrkl.ac.in

***Abstract*—The air pollution problem has become a critical issue worldwide due to its severe harmful impact on human health. Among all the pollutants, particulate matter particles are small enough to enter the lungs quickly and affecting the human body's respiratory system badly. Nowadays, air quality modeling has become an important research area to take supporting preventive action against the rising pollution level. However, air quality monitoring at every part of an area has been a complicated issue due to unstructured pollution monitoring sites. Much research has been conducted to predict air quality levels for pollution monitoring sites only, not for every part of a particular location, which is not useful for real-world scenarios. Therefore, this work proposes the CNN-GRU-RBF model based on the neural network approach, which considers both time and spatial features to solve this type of temporal-spatial interpolation issues of pollution prediction. The Convolutional Gated Recurrent Unit (CNN-GRU) is employed for feature learning and long-term temporal air quality modeling. Each site's time series prediction results are implemented as input for Radial Basis Function (RBF) spatial prediction layer to interpolate those time series prediction results, ultimately producing a better temporal-spatial interpolation map the study area. The proposed model shows its effectiveness during Spatio-temporal air quality prediction.**

***Index Terms*—Air quality, Spatio-temporal prediction, GRU, RBF**

I. Introduction

Ambient air pollution is one of the critical concerns for both developing and developed countries all over the world [1]. The increasing level of population growth and traffic congestion is the primary reason behind this critical issue. The deterioration of ambient air quality has a high negative impact on human health [2]. According to a few research studies, particulate matter pollutant is easy to inhale by causing a severe threat to the human body. According to the world health organization (WHO) report, the number of deaths is more than two million per annum in developing countries. Milion of fatalities occurs due to heart diseases, strokes, lung cancer, chronic respiratory problems, and many more related diseases. Many research analysis reveals the positive correlation among lung cancer, cardiovascular mortality with particulate matter concentration. Several research studies have noticed that ambient air quality levels in most cities do not meet the WHO guidelines for acceptable air quality. People of these cities have increased severe health risks related to the ambient air pollution level. The produced experimental results and analysis suggested a few public policies to control and mitigate ambient air pollution concentration.

Therefore, improving ambient air quality is one of the crucial steps in smart cities worldwide. Primarily, it will improve the poor health condition and also helps to gain economic growth significantly. In recent years, much research was conducted to perform air quality modeling to manage air quality in urban areas. Air quality modeling's primary goal is to predict the ambient air pollution level. Ultimately, it is used to ensure that the impact of air pollution levels on the human body should be minimal in the future.

Ambient air pollution concentration is not uniform in city areas, having a high value at smart cities, industrial estate, and traffic-congested roads. Its frequency varies over space and time due to several factors like meteorological variables and traffic. In various cities, many countries have developed air pollution monitoring sites. The number of monitoring stations in a town is not substantial due to the high cost of building monitoring sites. Therefore, air quality data should be perceived accurately before developing any air quality management system. Learning both spatial and temporal features gives better prediction results than training time and space dimensions separately. It arises the need of developing an efficient Spatio-temporal air pollution forecasting models.

There have been employed several Spatio-temporal air pollution forecasting models; still, the prediction results are not so accurate because most of the models predict air pollution levels for the existing monitoring sites only. Limited studies have conducted to predict air pollution level without monitoring stations, which arises missing values in space during Spatio-temporal air quality modeling. This is often called as temporal-spatial interpolation [3] issues of air quality predictions. It has observed that most of the traditional prediction models only focus on attributing the time series missing values, not missing values in space. These issues can be resolved by applying deep learning techniques by improving air pollution prediction results.

To overcome this problem, this research study works on both missing attributes in time series and space for a better air quality prediction results. This research paper employed deep learning and machine learning-based temporal-spatial interpolation methods to impute the missing value for a better Spatio-temporal forecasting model. This paper combines the

CNN layer [4] with GRU unit [5] to perform better temporal, spatial feature learning and time series modeling of pollutant value. The model then added an RBF interpolation layer, which imputes the missing values in space to perform spatial interpolation of data. Particulate matter PM_{10} is considered for the model evaluation purposes due to its severe negative impact on public health.

II. Related work

Time series based air quality modeling is a broad area in environmental research. Different types of statistical [6], machine learning, and deep learning research have been conducted based on temporal variations of air quality data. The statistical model includes an Autoregressive Moving Average (ARIMA) [7], Seasonal ARIMA (SARIMA) [8], Prophet [8] and the machine learning model comprises Decision Tree, Support Vector Machine (SVM) [9], Artificial Neural Network (ANN) [10], Adaptive Neuro-fuzzy Inference Systems (ANFIS) models, Multi-layer Perceptron, Principle Component Analysis (PCA) [11], Xgboost and Random Forest (RF) [12]. These models can provide better air quality prediction performance than traditional numerical air quality models. However, these models fail to handle the long term dependencies of the large historical dataset.

With the rapid development of artificial intelligence techniques, the concept of deep learning, an advanced variety of machine learning techniques evolved. These models are capable enough to train a large amount of input data for air quality modeling such as a Recurrent Neural Network (RNN) [13], Long Short Term Memory Network (LSTM) [14]–[17], Gated Recurrent Unit (GRU), Elman Neural Network, Temporal Convolutional Network (TCN) [18] and Time Delay Neural Network (TDNN). However, these shallow models consider either temporal features or spatial features at a time for air quality predictions. This research gap extends the further research study to develop a deep learning-based Spatio-temporal air pollution prediction model that can efficiently extract and learn both spatial and temporal features to improve prediction accuracy. Very few research was conducted to develop Spatio-temporal forecasting model [19] like Graph Convolutional Neural Network Long Short-Term Memory (GC-LSTM) [20], CNN-BILSTM [21] and Geo-LSTM [22]. Few of the spatial-temporal forecasting model uses satellite data as a source type to get better quality of prediction results. However, these models do not predict air pollution levels for each point in space, causing missing values in air pollution prediction results of a particular location. So, these models fail to predict the pollutant level where no monitoring stations are available. To remove these significant restrictions of air quality prediction, this research paper proposes a novel neural network-based temporal-spatial interpolation method that imputes the missing values in space and yields better Spatio-temporal prediction values.

Following the introduction, the rest of this research paper is organized as follows: Section 2 reviewed some existing related research work, and Section 3 describes the experimental study area of this work. Section 4 presents the proposed model architecture, named as CNN-GRU-RBF. Results and discussions are represented in Section 5, and the conclusion of the research is drawn in Section 6.

III. Study area

The study of this research paper is carried out at Odisha, an eastern part of Indian country. The state shares its boundaries with West Bengal, Chhattisgarh Jharkhand, and Andhra Pradesh spread over an area of 155,707 km^2. The state is blessed with several mineral reserves like coal, iron ore and chromite. This state is recorded as the second highest reserve of coal in the country [23]. Odisha is witnessed as the most top coal producer in the country. As per the environmental research report, coal extraction causes a severe negative impact on the human body due to the particulate matter generation during coal extraction. The neighborhood of coal mining areas witnesses a high concentration of air pollution. The coal mining process releases several toxic pollutants that adversely affect nearby locations. Traffic emission [24] air pollution is another primary source of toxic pollutants in Odisha. As per the National Clean Air Programme, Odisha has 6 six non-attainment cities that do not meet the air quality standard decided by Central Pollution Control Board. Therefore there is a requirement to analyze and predict the air quality level in Odisha so that necessary steps can be taken in advance against this critical situation. So Odisha is considered as the experiment location to perform temporal, spatial analysis [25] of air quality.

The primary data source considered for this study that comes from Odisha State Pollution Control Board [26], [27], having pollutant value for the duration 2004-2015. The experiments for this study are carried out over 16 air pollution monitoring stations of Odisha, having the most hazardous particle pollutant PM_{10} ($\mu\mathrm{g/m^3}$) concentration value and its geographical attributes.

IV. Proposed Method

A. Data preprocessing

The long term air quality data often face data missing issues due to sensor shutdown or some unusual activity. The missing attributes need to be replaced. It is necessary to have a pollutant concentration value for each day to analyze the time-varying air quality data. The proper analysis is conducted to find adequate interpolation techniques to impute those missing values for time series analysis. It is observed from the experimental study that linear interpolation provides better results than the other nonlinear interpolation methods. Therefore, this interpolation technique is employed to handle those missing values in the preprocessing step. It can be formulated as below [28],

$$\hat{x}(t) = \frac{x_{i+1} - x_i}{t_{i+1} - t_i}(t - t_i) + x_i \tag{1}$$

where $\hat{x}(t)$ is the linear interpolation function over time $t_{i+1} - t_i$.

Before performing the temporal and spatial analysis of the dataset, unnecessary outliers are removed and standardized using Z-Score normalization. The Z-score normalization can be formulated as below,

$$Z - Score_m = \frac{\sum_{m=1}^{n} p_m - \widehat{x}}{S} \tag{2}$$

where p_m presented the pollutant concentration value, $\widehat{x}$ and S represent mean and standard deviation respectively. In the process of training the model, 90 % of the dataset is utilized for training, 10 % for testing purposes. The proposed model has three stages, i.e., feature extraction layer (CNN layer), temporal analysis layer (GRU layer), and spatial analysis layer (RBF interpolation layer), as represented in Figure 1.

The detailed description of each step of the proposed architecture is presented in the following sections.

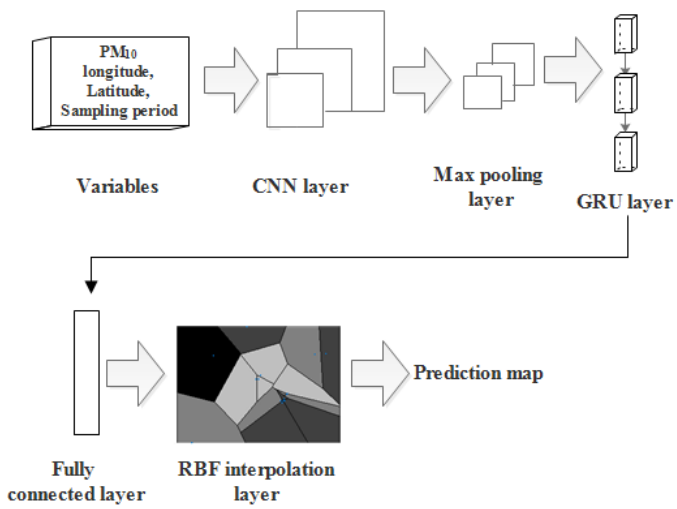

Fig. 1. The proposed model architecture

B. Convolutional neural network for feature learning

The normalized air quality dataset is utilized as the input for the Convolutional Neural Network (CNN) [29]. The CNN network has widely implemented image processing and computer vision applications, but we can use one-dimensional CNN for sequential modeling applications. The max-pooling operation of CNN reduces the dimensionality of the network and increases the model's speed and computational performance. Sparse connectivity and weight sharing are the key features of 1D CNN.

C. Long term time series modeling layer

In the case of a time series pollution dataset, as we need to deal with the longer sequential dataset, selective read, write and forget strategy should be followed to fetch the useful information only. Traditional Recurrent Neural Network architecture is not efficient enough to handle long term dependency and also suffer from vanishing gradient or exploding gradient issue in a sequence model. To handle this type of situation, GRU has applied to the output of the CNN layer, which overcomes the vanishing gradient issues, as well as analyzes the long term dependency in the air pollution dataset. It is simpler than Long Short Term Memory Neural Network (LSTM) and easy to implement. GRU is having reset gate and update gate where update gate works similar to the input gate and forget gate of LSTM model. GRU can be represented as [30],

$$\begin{aligned} z &= \sigma(W_Z x_t + U_z h_{t-1} + b_z) \\ r &= \sigma(W_r x_t + U_r h_{t-1} + b_z) \\ m &= \phi(W_m x_t + U_m (h_{t-1} \circ r) + b_m) \\ h_t &= (1 - z) h_{t-1} + z \circ m \end{aligned} \tag{3}$$

where, sigmoid function σ is used as activation function, x_t is used as the input at current time t, h_{t-1} is the past data at time $t-1$. (W_Z, U_z), (W_r, U_r), (W_m, U_m) are the weight parameters for update gate z, reset gate z and cell memory respectively. It combined the feature of the input gate and forgot the gate to form an update gate to make it simpler than LSTM. Update gate decides how much information should keep around and what to through away, while the reset gate decides the amount of information need to forget. The output of the GRU layer gives the prediction value of PM_{10} for each existing monitoring stations for the next 28 days (1-28 December 2015). Two-layer of GRU is used with 200 batch size, 1e-3 learning rate, 675 iterations, and 0.2 dropout value. It utilized Adam as an optimizer and Mean Square Error (MSE) as a loss function during the training process of the CNN-GRU model. The basic structure of GRU model is represented in Figure 2.

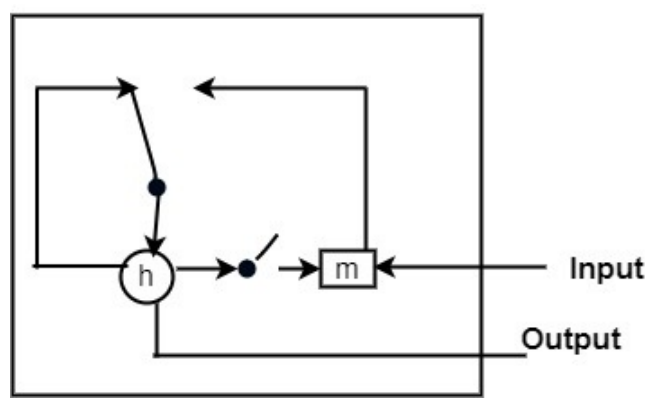

Fig. 2. Basic structure of GRU

D. Spatial interpolation layer

All the existing time series prediction models predict the pollution level at existing monitoring sites only. So it becomes challenging to perform air quality monitoring at unmeasured points [31]. It is also tough and cost-effective to construct monitoring sites in each part of a location. So it is essential to predict the spatial distribution of ambient air quality. To address this issue, we applied the RBF interpolation layer at the end of the CNN-GRU time series prediction layer so

that it can predict PM_{10} value at each corner of the study area at each time instance. Radial basis function (RBF) [32] interpolation has been used in many applications; it has a vital role in handling missing attributes in the geospatial dataset. To handle large computational tasks and to model complicated surfaces, it has shown its efficiency in many applications. This model is a series of exact interpolation techniques, having five basis kernel function, i.e., thin plate spline, spline with tension, completely regularized spline, multiquadric function, and inverse multiquadric function [33]. This is a function to compute the distance from each location in d-dimensional space. Unlike the inverse distance weighting technique, it is able to predict above the maximum of observed values and also below the minimum of the observed value. It can be utilized to generate a smooth surface from a large amount of dataset and can be computed mathematically as the weighted average of the data point value. It is basically based on the distance computation between two locations in d-dimensional space and can be represented by $f(x_0)$ function as below,

$$f(x_0) = \sum_{i=1}^{N} \lambda_i \varphi(\|x_0 - x_i\|) \tag{4}$$

where λ_i is the weight parameter, N represents the number of sampling point, φ represents a radial basis function, $(\|x_0 - x_i\|)$ represents the radial basis distance between the unknown point for which the new x_0 value is calculated and the measured point having known value x_i [34].

V. Results and Discussions

After training the CNN-GRU deep learning model, prediction value of PM_{10} for each spatial feature is obtained. The temporal modeling results show the prediction results for each monitoring station of the study area for the next 28 days, i.e., for December 2015. However, still, the prediction data are missing at the unmeasured points. Hence, the current paper performs both temporal modeling and spatial modeling to overcome this type of critical issues. Therefore, the RBF interpolation layer added at the top of the CNN-GRU model. The final results predict the PM_{10} concentration in the study area and for each geographical point where monitoring stations are not available and generate a temporal-spatial interpolation prediction map of the study area.

Figure 3-6 shows the prediction result of average PM_{10} value for every week in December 2015 in the study area. It can be seen from the weekly predicted map that the fourth week of December 2015 has more pollution levels than the other week, where the color scale indicates PM_{10} concentration level over the layer. A web Application is developed to display temporal-spatial interpolation map as shown in Figure 7.

The proposed CNN-GRU performance in temporal modeling is compared with the other state of the art neural network models like the GRU, LSTM and CNN-LSTM model, as shown in Table I. Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage

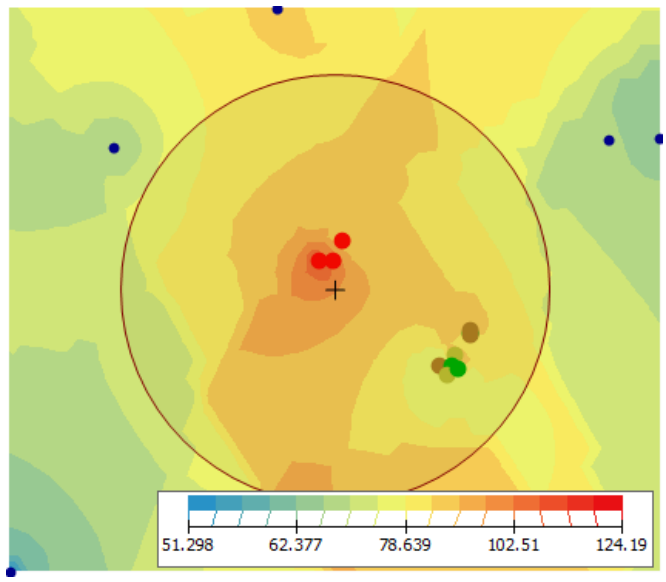

Fig. 3. Average predicted PM_{10} ($\mu g/m^3$) concentration distribution for first week of December 2015.

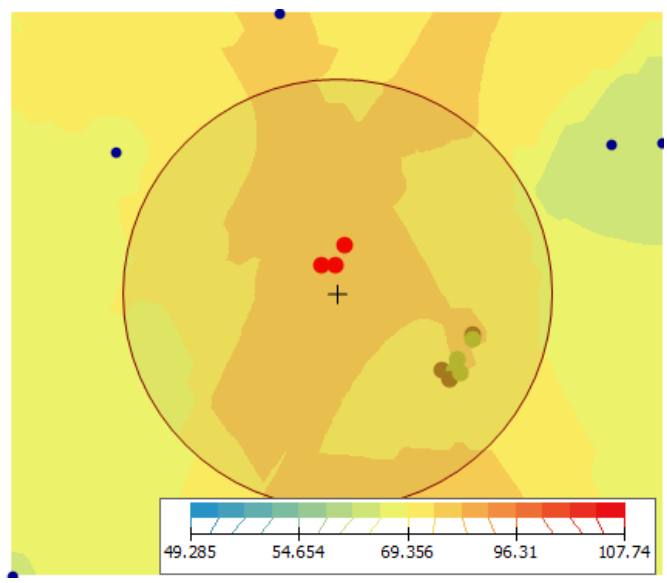

Fig. 4. Average predicted PM_{10} ($\mu g/m^3$) concentration distribution for second week of December 2015.

Error are utilized as error metrics to compare CNN-GRU's prediction performance in temporal modeling.

Table I shows that the CNN-GRU is a better performing prediction model due to lower RMSE, MAE, and MAPE values.

To verify the interpolation efficiency of the CNN-GRU-RBF model, Exponential Kriging (EK) [35], Universal Kriging (UK), Inverse Distance Weighting (IDW) [36] and Spherical Kriging (SK) models' interpolation performance conducted using CNN-GRU time-series prediction results for a fair comparison. Root Mean Square Error (RMSE) and Mean Error (ME) error metrics are utilized to evaluate the proposed

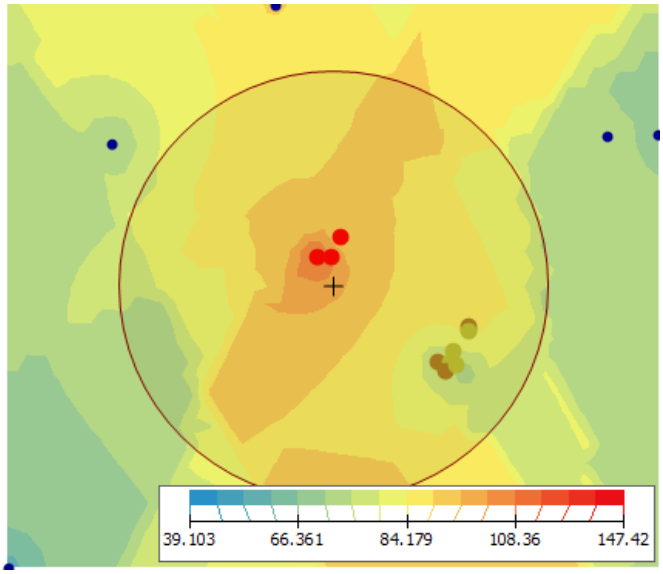

Fig. 5. Average predicted PM_{10} ($\mu g/m^3$) concentration distribution for third week of December 2015.

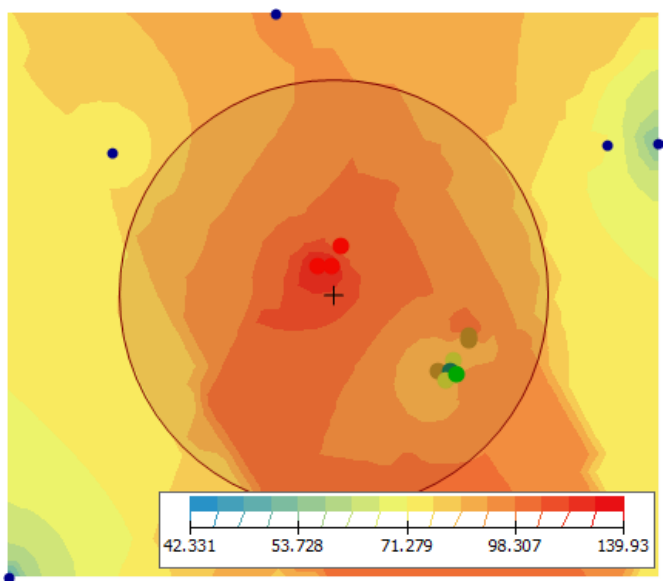

Fig. 6. Average predicted PM_{10} ($\mu g/m^3$) concentration distribution for fourth week of December 2015.

model's interpolation performance. The comparative analysis of these models is represented in Table II.

As shown in Table II, CNN-GRU-EK, CNN-GRU-UK, CNN-GRU-SK, CNN-GRU-IDW have higher RMSE and ME value as compared to the proposed model. In contrast, CNN-GRU-EK and the CNN-GRU-UK have similar performance. Furthermore, CNN-GRU-IDW performed well than the CNN-GRU-EK and CNN-GRU-UK methods. More significantly, the proposed model has superior performance in generating temporal-spatial interpolation maps as compared to other existing models. The key contribution of this research work can be summarized as follows,

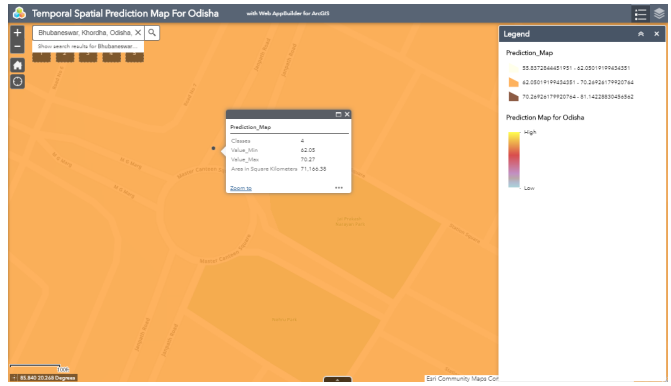

Fig. 7. Web Application

TABLE I
TIME SERIES PREDICTION PERFORMANCE COMPARISON OVER THE 28 DAYS

Model	RMSE	MAE	MAPE
GRU	56.43	46.43	38.48
CNN-LSTM	71.55	59.80	47.62
LSTM	57.24	49.16	37.51
CNN-GRU	53.40	40.03	31.30

- This is the first research experiment that integrated studies of deep learning techniques and machine learning-based interpolation techniques to solve the temporal-spatial interpolation issue of Spatio-temporal air quality modeling.
- The Traditional spatial prediction model imputes the missing values in the historical time series dataset. However, the proposed temporal-spatial interpolation model imputes the missing value both in the time domain and spatial domain to predict the pollutant concentration more accurately in advance at high temporal granularity.
- The CNN-GRU-RBF model can predict the long term exposure of PM_{10} concentration at each geographical point in the study area.

VI. CONCLUSION

In this work, spatial-temporal prediction experiments have conducted using the proposed CNN-GRU-RBF model. The CNN-GRU-RBF model utilizes both the neural network and the geostatistical concept to solve the temporal-spatial interpolation issues of the unmeasured point. The performance of the CNN-GRU-RBF model is investigated against the existing prediction models. The results show that the CNN-GRU-RBF model can solve the temporal-spatial interpolation issues

TABLE II
INTERPOLATION PERFORMANCE COMPARISON OVER THE 28 DAYS

Model	RMSE	ME
CNN-GRU-EK	25.60	4.05
CNN-GRU-UK	25.60	4.05
CNN-GRU-IDW	24.07	6.76
CNN-GRU-SK	23.67	4.00
CNN-GRU-RBF	22.30	2.86

more accurately than others. As compared to other models, the proposed method improves the prediction performance more significantly due to both temporal and spatial modeling abilities.

Due to the data unavailability, only PM_{10} concentration values are used for the period 2004-2015. The performance can be improved in the future by considering the correlation of pollution with the other affecting variables like traffic and meteorological factors.

VII. Acknowledgment

This research work was supported by the Ministry of Human Resource Development (MHRD) and Urban Development (UD) under the grant [Grant No.- 7794/2016].

References

[1] K Krishna Rani Samal, Korra Sathya Babu, and Santos Kumar Das. Ors: The optimal routing solution for smart city users. In *Electronic Systems and Intelligent Computing*, pages 177–186. Springer, 2020.

[2] Ying Chen, Oliver Wild, Luke Conibear, Liang Ran, Jianjun He, Lina Wang, and Yu Wang. Local characteristics of and exposure to fine particulate matter (pm2. 5) in four indian megacities. *Atmospheric Environment: X*, 5:100052, 2020.

[3] Mathieu Lepot, Jean-Baptiste Aubin, and François HLR Clemens. Interpolation in time series: An introductive overview of existing methods, their performance criteria and uncertainty assessment. *Water*, 9(10):796, 2017.

[4] Haofei Xie, Lin Ji, Quan Wang, and Zhejian Jia. Research of pm2. 5 prediction system based on cnns-gru in wuxi urban area. In *IOP Conference Series: Earth and Environmental Science*, volume 300, page 032073. IOP Publishing, 2019.

[5] Xinxing Zhou, Jianjun Xu, Ping Zeng, and Xiankai Meng. Air pollutant concentration prediction based on gru method. In *Journal of Physics: Conference Series*, volume 1168, page 032058. IOP Publishing, 2019.

[6] A Alimissis, K Philippopoulos, CG Tzanis, and D Deligiorgi. Spatial estimation of urban air pollution with the use of artificial neural network models. *Atmospheric environment*, 191:205–213, 2018.

[7] Shalini Jain and VL Mandowara. Study on particulate matter pollution in jaipur city. *International Journal of Applied Engineering Research*, 14(3):637–645, 2019.

[8] K Samal, Korra Sathya Babu, Santosh Kumar Das, and Abhirup Acharaya. Time series based air pollution forecasting using sarima and prophet model. In *Proceedings of the 2019 International Conference on Information Technology and Computer Communications*, pages 80–85. ACM, 2019.

[9] Nam-Uk Lee, Jae-Sung Shim, Yong-Wan Ju, and Seok-Cheon Park. Design and implementation of the sarima–svm time series analysis algorithm for the improvement of atmospheric environment forecast accuracy. *Soft Computing*, 22(13):4275–4281, 2018.

[10] Khaled Bashir Shaban, Abdullah Kadri, and Eman Rezk. Urban air pollution monitoring system with forecasting models. *IEEE Sensors Journal*, 16(8):2598–2606, 2016.

[11] Anikender Kumar and Pramila Goyal. Forecasting of air quality in delhi using principal component regression technique. *Atmospheric Pollution Research*, 2(4):436–444, 2011.

[12] Min Zhu, Jing Xia, Xiaoqing Jin, Molei Yan, Guolong Cai, Jing Yan, and Gangmin Ning. Class weights random forest algorithm for processing class imbalanced medical data. *IEEE Access*, 6:4641–4652, 2018.

[13] Hongqian Qin. Comparison of deep learning models on time series forecasting: a case study of dissolved oxygen prediction. *arXiv preprint arXiv:1911.08414*, 2019.

[14] K Krishna Rani Samal, Korra Sathya Babu, Ankit Kumar Panda, and Santos Kumar Das. Data driven multivariate air quality forecasting using dynamic fine tuning autoencoder layer. In *2020 IEEE 17th India Council International Conference (INDICON)*, pages 1–6. IEEE, 2020.

[15] K Krishna Rani Samal, Korra Sathya Babu, Abhirup Acharya, and Santos Kumar Das. Long term forecasting of ambient air quality using deep learning approach. In *2020 IEEE 17th India Council International Conference (INDICON)*, pages 1–6. IEEE, 2020.

[16] K. K. R. Samal, K. S. Babu, and S. Das. Spatio-temporal prediction of air quality using distance based interpolation and deep learning techniques. 2021.

[17] K Krishna Rani Samal, Ankit Kumar Panda, Korra Sathya Babu, and Santos Kumar Das. An improved pollution forecasting model with meteorological impact using multiple imputation and fine-tuning approach. *Sustainable Cities and Society*, page 102923, 2021.

[18] K Krishna Rani Samal, Korra Sathya Babu, and Santos Kumar Das. Multi-directional temporal convolutional artificial neural network for pm2. 5 forecasting with missing values: A deep learning approach. *Urban Climate*, 36:100800, 2021.

[19] Reza Asadi and Amelia Regan. A spatial-temporal decomposition based deep neural network for time series forecasting. *arXiv preprint arXiv:1902.00636*, 2019.

[20] Yanlin Qi, Qi Li, Hamed Karimian, and Di Liu. A hybrid model for spatiotemporal forecasting of pm2. 5 based on graph convolutional neural network and long short-term memory. *Science of the Total Environment*, 664:1–10, 2019.

[21] Shengdong Du, Tianrui Li, Yan Yang, and Shi-Jinn Horng. Deep air quality forecasting using hybrid deep learning framework. *arXiv preprint arXiv:1812.04783*, 2018.

[22] Jun Ma, Yuexiong Ding, Jack CP Cheng, Feifeng Jiang, and Zhiwei Wan. 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, 2019.

[23] Tapaswini Nayak and Indrani Roy Chowdhury. Health damages from air pollution: Evidence from open cast coal mining region of odisha, india. *Ecology*, 1(1):42–66, 2018.

[24] Anu Rani Sharma, Shailesh Kumar Kharol, and KVS Badarinath. Influence of vehicular traffic on urban air quality–a case study of hyderabad, india. *Transportation Research Part D: Transport and Environment*, 15(3):154–159, 2010.

[25] Jannah Baker, Nicole White, and Kerrie Mengersen. Missing in space: an evaluation of imputation methods for missing data in spatial analysis of risk factors for type ii diabetes. *International journal of health geographics*, 13(1):47, 2014.

[26] OpenGovermentDataPlatformIndia. Ambient air quality data of odisha, 2017, Oct 16.

[27] CentralPollutionControlBoard. Air pollution, 2018, Jan 6.

[28] Albrecht Gnauck. Interpolation and approximation of water quality time series and process identification. *Analytical and bioanalytical chemistry*, 380(3):484–492, 2004.

[29] Bendong Zhao, Huanzhang Lu, Shangfeng Chen, Junliang Liu, and Dongya Wu. Convolutional neural networks for time series classification. *Journal of Systems Engineering and Electronics*, 28(1):162–169, 2017.

[30] Jaehyun Ahn, Dongil Shin, Kyuho Kim, and Jihoon Yang. Indoor air quality analysis using deep learning with sensor data. *Sensors*, 17(11):2476, 2017.

[31] Phuong T Vu, Timothy V Larson, and Adam A Szpiro. Probabilistic predictive principal component analysis for spatially misaligned and high-dimensional air pollution data with missing observations. *Environmetrics*, page e2614, 2019.

[32] Vaclav Skala. Rbf interpolation with csrbf of large data sets. *Procedia Computer Science*, 108:2433–2437, 2017.

[33] Pengwei Qiao, Peizhong Li, Yanjun Cheng, Wenxia Wei, Sucai Yang, Mei Lei, and Tongbin Chen. Comparison of common spatial interpolation methods for analyzing pollutant spatial distributions at contaminated sites. *Environmental geochemistry and health*, pages 1–22, 2019.

[34] Maxim Arseni, Mirela Voiculescu, Lucian Puiu Georgescu, Catalina Iticescu, and Adrian Rosu. Testing different interpolation methods based on single beam echosounder river surveying. case study: Siret river. *ISPRS International Journal of Geo-Information*, 8(11):507, 2019.

[35] SB Choi, Geneveve Parreno, Kyu Kon Kim, and CW Kang. A study on imputation for missing data using the kriging. *Journal of the Korean Data Analysis Society*, 17(6):2857–2866, 2015.

[36] Mohammed M Shareef, Tahir Husain, and Badr Alharbi. Optimization of air quality monitoring network using gis based interpolation techniques. *Journal of Environmental Protection*, 7(6):895–911, 2016.