Journal of Universal Computer Science, vol. 28, no. 8 (2022), 799-826
submitted: 5/12/2021, accepted: 7/7/2022, appeared: 28/8/2022

A Deep Learning based Air Quality Prediction Technique Using Influencing Pollutants of Neighboring Locations in Smart City

Banani Ghose
(Haldia Institute of Technology, Haldia, West Bengal, India
https://orcid.org/0000-0001-6558-1334, bananighose@hithaldia.in)

Zeenat Rehena*
(Aliah University, Kolkata, West Bengal, India
https://orcid.org/0000-0002-0689-1903, zeenatrehena@yahoo.co.in)

Leonidas Anthopoulos
(University of Thessaly, Greece
https://orcid.org/0000-0002-7731-4716, lanthopo@uth.gr)

Abstract: The level of air pollution in smart cities plays a critical role in the community's health and quality of life. Thus, air pollution forecasting would be beneficial and would guide citizens in avoiding exposure to dangerous emissions. The air health of a place can be diagnosed by close observation of the AQI (Air Quality Index) of that place. Moreover, the AQI of a place may have some influence on the pollutant concentration of the neighboring places. To address this issue, this work introduces a hybrid deep learning framework that is able to predict the values of a corresponding metric: AQI of smart cities. As a part of this work, two algorithms are proposed. The first one replaces the missing values in the dataset and the second one formulates the influence of the nearby places' pollutant concentrations on the air quality of a particular place. A deep learning-based forecasting model is also proposed by combining 1D-CNN and Bi-GRU. To test the applicability of the framework, a large-scale experiment is carried out with the real-world dataset collected from New South Wales, Australia. Experimental results validate that the proposed framework provides a stable forecasting result, it confirms that the AQI of a place gets affected by the pollutant concentration of the nearby places and the comparison of forecasting result with the existing state of the art models shows that the proposed model outperforms the other models.

Keywords: Smart city, Air quality, Prediction, Atmosphere, Deep learning
Categories: H.3.1, H.3.2, H.3.3, H.3.7, H.5.1
DOI: 10.3897/jucs.78884

1 Introduction

Modern urbanization and the smart city concept resulted in major dependency on advanced technologies to guarantee the quality of life to the citizens [Rouse, 2019]. There are several challenges that are to be faced every moment by the residents of smart cities. Some of them are like water pollution, lowering in water levels, polluted environment, noise pollution, security issues, and many more. Smart cities are the worst sufferers of air pollution which are caused by unrestricted industrialization as well as the increase in the means of transportation means. To get rid of this worsening environmental condition, one may prefer to avoid the severely polluted zones in the cities and plan their way out. The policymakers may also prefer to have a hint of pollution beforehand so that they can take necessary steps that will guide the citizens to avoid the polluted areas in the cities. Predicting air pollution is thus a very important task. AQI is a standard indicator for assessing the air health of a place at a specific time. Many researchers are proposing many ways to forecast air pollution from a smart city perspective.

With the advancement of technologies, two efficient approaches came in. Those are machine learning and deep learning approaches. Some of the popular ML approaches like SVR (Support Vector Regressor), ARIMA (Auto-Regressive Integrated Moving Average) have delivered their remarkable contributions to air pollution forecasting. For the successful deployment of deep learning architectures a large volume of the historical dataset is necessary. Recent up-gradation of technology has allowed having these kinds of datasets with the help of IoT and Big data, especially in smart cities where much-developed technologies and networks are deployed for different purposes for the betterment of smart cities. Deep learning caters to more accurate prediction results than machine learning [Gamboa, 2017]. Among all deep learning constructs, RNN (Recurrent Neural Network) had gained popularity as it had excellent time-series data handling capabilities and has a better ability in temporal feature extraction. Besides many interesting features, RNN suffers from vanishing gradient problems and exploding gradient problems. RNN cannot address the long-term dependencies, which is very important in prediction-related works. To subside the problems of RNN, two variations of RNN were deployed. They are GRU (Gated Recurrent Unit) and LSTM (Long Short Term Memory). These two constructs have memory units and they are experts in handling long-term dependencies. GRU also exhibits better prediction results [Gao et al., 2019]. In recent days another deep learning architecture came into practice, named CNN (Convolution Neural Network). CNN has remarkable success in image classification [Wei et al., 2015]. CNN has automatic feature extraction as well as parameter sharing capacity. One of the variations of CNN is 1D-CNN (1 Dimensional Convolution Neural Network). 1D-CNN has shown satisfying results in time-series analysis [Du et al., 2021]. In the last three years, hybrid models have been used where several deep learning architectures have been deployed. The resultant model enjoys the benefits of its basic building blocks. The hybrid architectures exhibit better prediction accuracy [Du et al., 2021, Tao et al., 2019, Zhu et al., 2017]. Reviewing the related works, it is observed that several challenges lie in these fields which have not yet been considered. Most of the works concentrate on the pollutant concentration and/or meteorological factors of a given place to forecast the AQI of that place. The neighboring places' pollutant concentration and meteorological factors' influences are not addressed yet, though it is possible that these are related. The dataset bearing several missing values is to be properly managed to get a more perfect prediction result, which is another area to be considered a challenge. In this regard, this paper focuses on the prediction of the air pollution of a place in a smart city, based on the above factors. More specifically, it aims to provide answers to the following research questions:

1. RQ1: Is the air quality of a place affected by the pollutant concentration and meteorological factors of the nearby places?

2. RQ2: How can deep learning methods are utilized in predicting corresponding Air Quality Index values?

Both these questions need to be answered because community health is affected by air quality factors [Che et al., 2020, Leng et al., 2020]. In [Rehena and Janssen, 2019] the promises of the smart city infrastructures are listed. While in [Schürholz et al., 2020], similar problems related to smart cities are elaborated upon. For searching the answer of RQ1, an algorithm is designed (in Section 3.5) and it is tested through proper experimentation with a positive outcome (Section 5.3). Concerning RQ2, an advanced hybrid deep learning-based air pollution forecasting framework is proposed. The framework is called "Deep Learning-based Air Quality Prediction Framework (DLAQPF)". Initially, DLAQPF constructs a combined dataset from a local station and its neighboring stations. This combined dataset is then fed into the hybrid prediction model which comprises a stack of 1D-CNN and a stack of Bi-directional Gated Recurrent Unit (Bi-GRU). There are several facts responsible for the decision to choose GRU and CNN as the primary constructs. GRU requires less training time because it has fewer gates. Bi-GRU consists of two ordinary GRUs. This unit can address both way dependencies in time-series data. 1D-CNN performs better in automatic feature extraction without human interference. It avoids pre-training which speeds up the whole training process. 1-D CNN is also capable of parameter sharing. 1D-CNN has an outstanding dimension reduction capability.

In the proposed work, the following key contributions are done:

1. Air quality prediction framework is proposed for a smart city.
2. An imputation algorithm is designed to deal with the missing values.
3. Dataset dimension is reduced through correlation analysis and an algorithm to determine the influence of pollutant concentration of neighboring places for predicting the AQI of a place is framed.
4. As a part of the air quality prediction framework, a hybrid deep learning model is developed by combining two deep learning architectures, such as stacked 1D-CNN and stacked Bi-GRU.

The rest of this paper is organized as follows: Section 2 contains some background surveys carried out to formulate the task altogether. Section 3 contains the research methodology of the proposed work that concludes with the definition of the framework. Section 4 demonstrates an experiment to validate the applicability of the framework. In Section 5, the results which are yielded after experimentation are explained and briefly discussed. Finally, Section 6 contains conclusions and some future thoughts.

2 Literature Review

Predicting air pollution in smart ensures the good health of its citizens. It helps in policymaking and individual decision-making. Several researchers are working in this field. Broadly, the methods can be categorized into two approaches namely: non-machine learning approaches which include a statistical or numerical method, and machine learning approaches. The statistical-numerical method is a good old method as used in [Lee et al., 2015, Vardoulakis et al., 2003, Zhang et al., 2012]. With the advancement of computational power and emerging IoT technology in smart cities, a large volume of sensor data is easily available. It motivates the researchers to deploy deep learning approaches in time-series forecasting problems. In recent times numerous machine learning [Martínez-España et al., 2018] and deep learning-based research works are done for predicting air pollution [Bekkar et al., 2021]. Many researchers have proposed some forecasting techniques using trending technology like IoT [Minoli et al., 2017]. [Wu and Lin, 2019] have framed an AQI forecasting technique by deploying LSTM in it. In [Liu et al., 2021b, Wang et al., 2019], the historical smog data are used to forecast the smoke probability using an algorithm based on LSTM and GRU. In, [Yang et al., 2016] the researchers used the wavelet analysis method to forecast air pollution concentration in [Liu et al., 2021a]. Wavelet analysis method is also used in [Yang et al., 2016] to decompose SO_2, finally using the Fourier curve, a better prediction result is accomplished. In [Qianrao, 2016] and [Donnelly et al., 2015] the authors preferred regression techniques for efficient forecasting of haze and air pollution respectively. With the advancement in technology, the deep learning concept has already proved its efficiency in many research works. A weather prediction model was proposed by [Ren et al., 2021] using deep learning constructs. A comparative study among the performances of various deep learning approaches to predict air pollution in smart cities is formulated in [Ameer et al., 2019]. In [Ghose and Rehena, 2021], a Deep Long Short Term Memory Model (DLSTM) is proposed to predict the air quality where the authors considered the pollutant concentrations of a place to predict the AQI of that place. A hybrid deep learning model is suggested in [Du et al., 2021] and to predict air pollution Deep Air Quality Forecasting Framework (DAQFF) model is proposed. In this paper, the air quality data is highly dynamic and its non-linear nature is focused upon. In this paper, the authors considered the spatial-temporal behaviour of air quality data. CO_2 is one of the major pollutants. CO_2 concentration is predicted using a machine learning approach in [Deleawe et al., 2010]. In [Yi et al., 2018], the researchers proposed methods to forecast air pollution using big data concepts. In [Verma et al., 2018] the air quality data in every 6th, 12th, and 24th hours are captured from the sensors and air pollution is predicted using another deep learning model named Bi-directional LSTM. In [Kurt et al., 2008], the authors proposed a neural network-based solution to predict the concentration of three major air pollutants like SO_2, PM10, and CO for the next three days. In [Tao et al., 2019], the Convolution-based Bi-directional Gated Recurrent Unit (CBGRU) model is proposed using a combination of 1D-CNN and Bi-GRU to predict the PM2.5 using the Beijing dataset. [Zhu et al., 2017] introduced a hybrid air pollution forecasting model, which is also deployed on the Beijing dataset.

By reviewing the existing works, it has been observed that different researchers formulated their works using statistical or numerical techniques in earlier days, later, the researchers concentrated on machine learning, deep learning, and hybrid deep learning architectures to predict the air quality of a particular place more accurately. Whatever architectures they had used, they focused on predicting the air quality depending upon the most influencing pollutant agent or by considering the pollutant concentration and/or the meteorological factors of the place for which they wanted to predict the air health. But, none of them considered the effect of the concentration of pollutants in the nearby places to forecast the AQI of a specific place. In this paper, to predict air pollution in a locality, a unique framework named DLAQPF is proposed. The framework focuses on the dataset comprising meteorological data as well as pollutant concentrations of a particular place and its nearby places. At the same time, it formulates the influence of the pollutants of nearby places to predict the AQI of that place. As part of this framework, a hybrid forecasting model is proposed by combining

1D-CNN and Bi-GRU deep learning architectures. This model is influenced by the model CBGRU proposed in [Tao et al., 2019]. In comparison with CBGRU, in the CNN block of the proposed hybrid forecasting model, instead of one 1D-CNN layer, a series of three 1D-CNN layers are deployed to extract more complex local features and a drop out layer is introduced to deal with the overfitting problem. In the GRU block also three layers of Bi-GRU are used to learn both ways long-term temporal dependencies instead of two layers of Bi-GRU.

3 Proposed Methodology

In this section, the detailed methodology of the proposed work is described. First, the motivation and problem statement of the work is explained followed by other subsections. In the subsequent subsections the overview of the proposed work is discussed, later the missing value problem is resolved, next, as a part of the work the correlation analysis is carried out, the influence of the neighboring stations is evaluated and finally the hybrid prediction model is explained in details.

3.1 Problem Statement & Motivation

Predicting air pollution in a smart city is a challenge nowadays. Air pollution forecasting leads to ensuring good environmental ambiance and thus provides better health to the citizens. To forecast air pollution, a strong dataset is needed. The dataset consists of the observations related to pollutants and the meteorological factors in the air. The observation is recorded as per the data collected by the sensors installed in the environment. Usually, the data comprises hourly observations. The concentrations of pollutants like PM2.5, PM10, SO_2, NO_2, NO, CO_2, O_3, C_6H_6, NH_3 are predominantly recorded. Temperature, humidity, wind speed, wind direction, solar radiation, snowfall, rainfall are also kept in the record. Thus, the pollution-related dataset is multivariate time-series data. The air quality of a place can be ascertained by taking these factors into consideration. The existing works have addressed the prediction-related problem efficiently, but some challenges are yet to be addressed. The previous works were centred on the AQI prediction by focusing on the pollutant concentration and meteorological factors of that place only. It is a challenge to address the influence of the surrounding places' pollutant concentration in predicting the AQI of a place. Another challenge is to consider the temporal both-way dependencies of the time-series data. The main objective of this work is to develop a robust deep learning-based forecasting model that will be able to predict the air health of a place more accurately by considering the pollutant concentrations of the nearby places.

3.2 Overview of the proposed work

In this study, a deep learning-based air quality prediction framework (DLAQPF) has been proposed. The flowchart of the DLAQPF is represented in Figure 1. This framework consists of four phases, such as missing data imputation, neighboring places' impact determination, correlation analysis, and deep learning model construction. In the missing data imputation phase, a seasonality-based imputation algorithm is proposed to impute the missing values more accurately. In the next phase, the impact of nearby places' PM2.5 concentration is determined to predict the AQI of a place. In the correlation analysis phase, the correlation between the pollutant concentrations and the meteorological factors is calculated to refine the dataset without losing its impact on AQI calculation. Lastly, a hybrid deep learning-based model is constructed by combining 1D-CNN and Bi-directional Gated Recurrent units (Bi-GRU) architectures to predict the AQI.

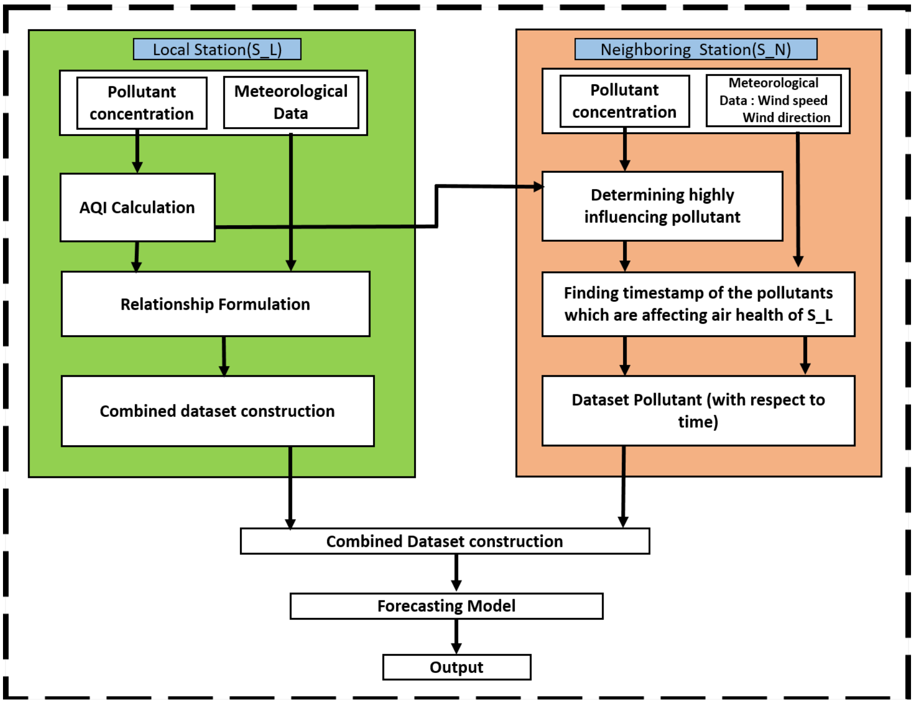


Figure 1: Flowchart of the proposed deep leaning based air quality prediction frame work.

In this work, the following dataset is used. The description of it is provided below.

– **Dataset Description:** In this work, six places in New South Wales, Australia are considered depending on the availability of the dataset, as depicted in Figure 2. Among these six places, one is selected as the local station and denoted by S_L, and the five other stations are taken as neighboring stations of S_L and are denoted by S_Ns. Here, Rozelle is considered to be S_L. Richmond, St. Marrys, Bringelly, Liverpool, and Randwick are the neighboring stations considered as S_Ns. The dataset from these six places are collected individually and they consist of two types of parameters:

– Pollutant concentrations.
– Meteorological factors.

The dataset is consisting of the concentration of PM10(pphm), PM2.5(pphm), CO(pphm), NO(pphm), NO_2(pphm), SO_2(pphm), and O_3(pphm) and meteorological factors like wind speed(m/s), rainfall(mm/m2), wind direction(°), humidity(%),, and temperature(°C). The dataset bears data from 01/01/2018 01:00 pm to 31/12/2020

01:00 pm. The total number of data operated here is exactly 26,280. It consists of data on an hourly basis. The dataset is collected from [NSW, 2022].

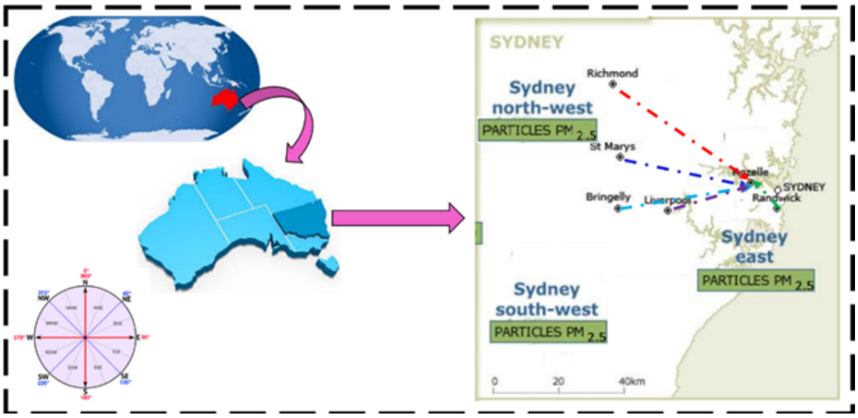


Figure 2: Location of data capturing stations. [AustraliaNewSouthWales, 2022, Direction, 2022, Worldmap, 2022]

The time-series dataset consisting of the pollutant concentration and meteorological factors (from 20.02.2018 12:00 hours to 06.05.2018 11:00 a.m.) are plotted in Figure 3. By a close observation of each of the graphs, it is noted that the time-series data is repeating its pattern in a fixed time interval. From this, it can be ascertained that the time-series data has some kinds of seasonality, namely:

– Daily seasonality
– Weekly seasonality

The air quality factors, if minutely observed in Figure 3, are repeating their pattern on a daily basis. It means every day the value of air health related factors remains almost the same at specific time stamps. Moreover, it is also observed that the pattern on a particular day's particular hour is being repeated the same day's specific hours in every week. The autocorrelation coefficient of the air quality factors also confirms the seasonality factor of the time-series dataset.

3.3 Missing Value Replacement

While collecting and preparing the dataset, it is observed that some data values are missing. Missing values are those values, which are not present in the dataset. The missing value problem may be caused due to many reasons like inappropriate method of data collection or mistakes committed at the time of storing the data, or the inability/failure of one or more sensors to capture data. Whatever be the cause, the consequences have a severe impact on the expected output. Here, the dataset is also suffering from a missing value problem. After analyzing the dataset, it is seen that there are two types of problems. One, the data is absent for a particular hour, and second, the data values are absent for more than one consecutive hours. There are numerous methods for missing value replacement. Here, a seasonality-based imputation algorithm is proposed to replace the missing values within the dataset and as described in Algorithm 1. Here, from statistical analysis of the dataset, it is observed that there are weekly and daily seasonality in the dataset as shown in Figure 3. So the data of a particular time of a day is most likely to be the same as the data of the previous week's same day same hour, and also the next week's same day and time. That means the data of a particular hour is in all probability to be the same as the data of 168th hour before and after. So, in this algorithm, the seasonality parameter θ is considered as 168 ($24 \times 7 = 168$). Here for the proposed seasonality-based imputation algorithm, this concept of seasonality is applied. In this proposed algorithm, two types of missing values are handled viz. single hour missing value and consecutive hours' missing value. If a value of a time-series Di at the timestamp j is missing, and the value for the timestamp (j - 1) and (j + 1) is present then the imputed value DMi,j is calculated by using the following formula:

$$DM_{i,j} = \frac{D_{i,j-1} + D_{i,j+1}}{2} \tag{1}$$

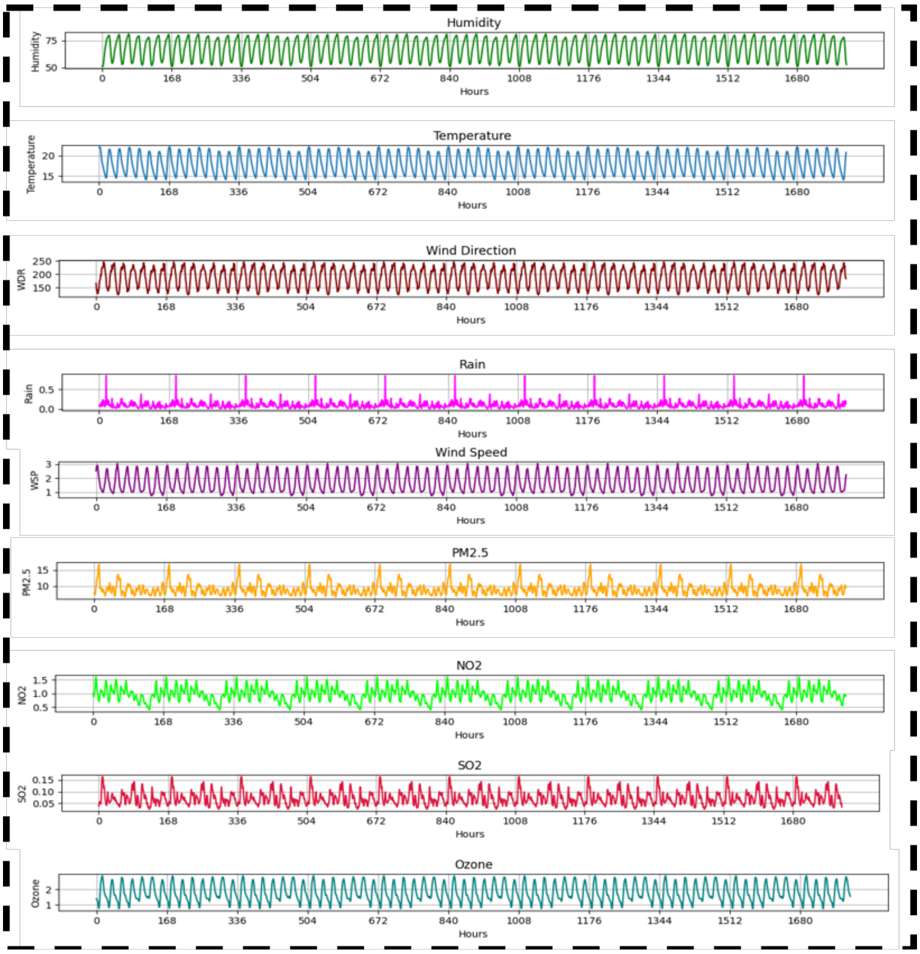


Figure 3: Graphical representation of each of the pollutant and meteorological factors from 20.02.2018 12:00hours to 06.05.2018 11:00a.m.

On the other hand, if a value of the time-series D_i is missing for the timestamp j and the value $D_{i,j+\eta}$ (η=1, 2, …, ν) is also missing, where $\nu \geq 1$, the imputed value $DM_{i,j}$ is determined by using the following formula:

$$Mean_1 = \frac{1}{\delta_1} D_{i,p} \quad ; \forall D_{i.p} \neq \text{null}, (p= (j-168), (j-2\times168),\dots, (j-s_1\times168)) \qquad (2)$$

$$Mean_2 = \frac{1}{\delta_2} D_{i,q} \quad ; \forall D_{i.q} \neq \text{null}, (p= (j+168), (j+2\times168), \dots, (j+s_2\times168)) \qquad (3)$$

$$DM_{i,j} = \frac{1}{2}(Mean_1 + Mean_2) \qquad (4)$$

Here δ_1 is the total number of non-missing data ($D_{i,p}$) and (j - s_1 ×168) > 0. δ_2 represents the total number of non-missing data ($D_{i,q}$) and (j +s_2 ×168) ≤ N. Table 1 shows the notations that are used in Algorithm 1.

Symbol	Meaning
D_{ij}	jth. Data of the D_ith. time-series
DM_{ij}	jth. Data of the DM_ith time-series
N	Number of input time-series
M	Number of data points in each time-series
θ	Seasonality parameter

Table 1: Notation used in Algorithm 1.

Algorithm 1 Seasonality-based imputation algorithm

Input: time-series $D_1, D_2, \dots, \dots, D_N$
Output: Modified time-series DM_1, DM_2, …, …, DM_N

i←1
while i ≤ N **do**
  for j←1 to M **do**
    $DM_{i,j} \leftarrow D_{i,j}$
    if (j≠1 and j ≠ N) and ($D_{i,j}$ = NULL and $D_{i,j-1}$ ≠ NULL and $D_{i,j+1}$ ≠ NULL) **then**
      $DM_{i,j} \leftarrow (D_{i,j-1} + D_{i,j+1}) / 2$
    else if $D_{i,j}$ = NULL and $D_{i,j+1}$ = NULL **then**
      k ← j - θ
      count ←0
      sum ← 0
      while k > 0 **do**
        sum ← sum + $D_{i,k}$
        count ←count + 1
        k ← k – θ
      end while
      k ← j + θ
      while k ≤ N **do**
        if $D_{i,k}$ ≠NULL **then**
          sum ← sum + $D_{i,k}$
          count ← count +1

```
                end if
                k ← k + θ
            end while
            DM_{i,j} = sum / count
        end if
    end for
    i ← i + 1
end while
```

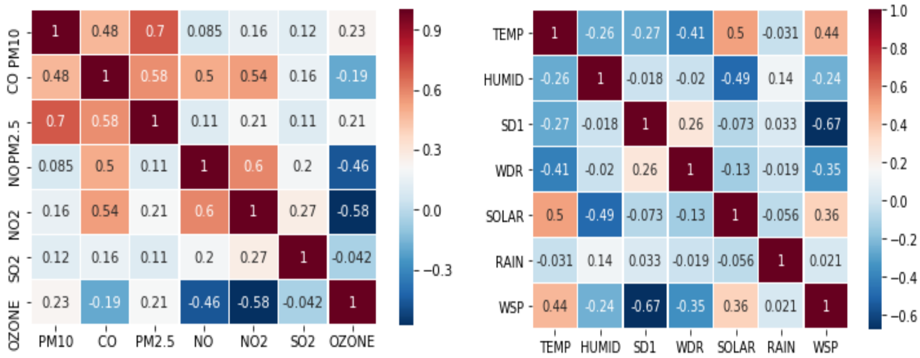
3.4 Correlation Analysis

To work with multivariate time-series data the complexity of the work becomes higher when the number of parameters is large. To reduce the complexity of the input and discard the undesirable overheads it is important to concentrate on the factors which are influencing the AQI of that place [Ghose and Rehena, 2020]. For reducing the dataset, a correlation among the pollutants is established. Later the same is done with the meteorological factors also. To calculate the correlation the following formulae is used:

$$\boldsymbol{R} = \frac{(N\sum AB - (\sum A \sum B))}{\sqrt{(N\sum A^2 - (\sum A)^2)(N\sum B^2 - (\sum B)^2)}} \tag{5}$$

Where, N= Number of pairs, A and B are individual sample points.

To get the refined dataset, the correlation among the pollutants is calculated as shown in Figure 4a. The highly correlated pairs are considered and one of the highly correlated pollutants is discarded. A close look at the correlation matrix in Figure 4a shows PM10 is highly correlated with PM2.5 and CO. From this set, PM2.5 is chosen and PM10 and CO are discarded. Similarly, NO_2 is highly correlated to NO, so from this pair, only NO_2 is chosen. SO_2 and O_3 are pollutants that are not remarkably correlated to any other pollutants. So, both of them are taken into the effective dataset. Ultimately, the refined dataset consists of pollutants like PM2.5, NO_2, SO_2, and O_3. In the same way, the set of meteorological factors are focused on to frame the concise dataset. For this, the correlation among the meteorological factors is calculated as shown in Figure 4b. As the matrix shows, TEMP and SOLAR are highly correlated, so discarding the SOLAR, only TEMP is taken in the refined dataset. There is a high-level correlation between SD1 and WSP, so leaving the SD1 behind, the WSP is considered. HUMID, RAIN, and WDR are not correlated with the other meteorological factors, so all of them are included in the dataset. Finally, the concise meteorological dataset comprises TEMP, HUMID, WDR, RAIN, and WSP.



(a) *(b)*

Figure 4: (a) Correlation among pollutants, (b) Correlation among meteorological factors.

Both the pollutant concentration and meteorological factors are considered to predict the AQI of the S_L. The AQI of the S_L is calculated considering the following formula as mentioned in [USEPA, 2020]

$$SIp = \left[\frac{(SI_{HI}-SI_{LO})}{(BP_{HI}-BP_{LO})} * (CO_P - BP_{LO})\right] + SI_{LO} \qquad (6)$$

Here, CO_P= the concentration of the pollutant p.
SI_P = the Sub-index of a specific pollutant concentration CO_P.
BP_{HI} = the Breakpoint concentration greater than or equal to the given concentration.
BP_{LO} = the Breakpoint concentration less than or equal to the given concentration.
SI_{HI} = the AQI value corresponding to BP_{HI}.
SI_{LO} = the AQI value corresponding to BP_{LO}.
Finally,

$$AQI = Max(SI_P) \qquad (7)$$

Where, p=1, 2, 3,..., n; denoting n pollutants.
Here, the sub-index of a specific pollutant concentration CO_P is calculated according to the formula mentioned in Equation 6. After calculating the sub-indices of all the concerned pollutant concentrations, the AQI of a particular place is determined by finding out the maximum sub-index calculated for a specific pollutant concentration; as mentioned in Equation 7.

Here to find the impact of the historical AQI value on the current AQI value, the autocorrelation function (ACF) is evaluated on the AQI time-series. If the correlation is determined to measure the interrelationship between lagged values of a time-series itself, it is called autocorrelation. For a time-series X_i(i = 1, 2, …, N), the autocorrelation function for the lag period l is defined as follows:

$$\lambda_l = \frac{\operatorname{cov}(X_i, X_{i-l})}{\sqrt{var(X_i)var(X_{i-l})}} \qquad (8)$$

Where, l = a, 2, …, N,

Here, l and N represent the lag period and length of the time-series respectively. The $cov(X_i,X_{i-l})$ denotes the covariance of X_i and X_{i-l}. The $var(X_i)$ and $var(X_{i-l})$ represent the variance of X_i and X_{i-l} respectively.

The autocorrelation graph of the AQI time-series for the lag period of l (l = 1, 2, …, 30) hours is shown in Figure 5. From this figure, it can be seen that there is a constant decrease in ACF up to 19 hours lag, then after it starts increasing up to 24 hours lag. From the 25th hour lag, it again starts decreasing up to 30 hours lag. For this, it can be concluded that there are a 24 hours seasonality pattern exists in the AQI time-series. So, in this study, to predict the (i + 1)th hour AQI current 24 hours (X_{i-23}; X_{i-22}; …, X_i) data is considered.

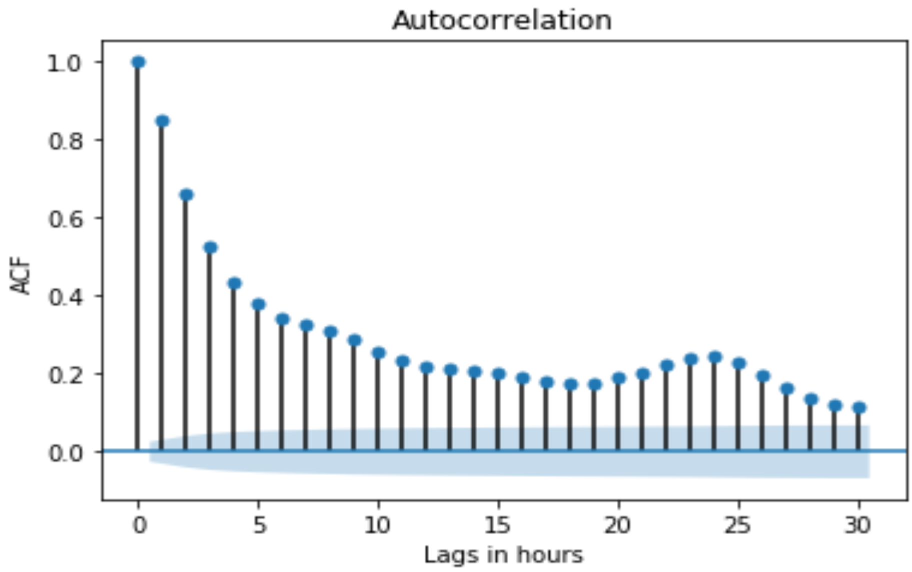


Figure 5: Autocorrelation function graph of the AQI time-series for 30 hours lag.

3.5 Evaluating the Influence of Neighboring Areas

In the recent research work [Liu et al., 2019], it has already established that AQI of a particular place indicates the air health of a specific place. In this letter, whether there is an influence of the pollutant concentration and meteorological factors on the AQI of the neighboring places (S_Ns) has been investigated. To validate this fact, here a neighboring place's impact (NPI) algorithm is proposed, which is presented in Algorithm 2. This algorithm will help in showing the influence of the air of S_Ns on predicting the AQI of S_L. Among the pollutants of S_Ns, here PM2.5 has been considered to calculate the AQI of S_L, as PM2.5 is the highest influencing factor to calculate AQI [Honarvar and Sami, 2019, Kim et al., 2015]. The notations which are used in the algorithm are shown in Table 2.

This algorithm generates a multivariate time-series datasets Y_i with two parameters such as a timestamp (T) and a PM2.5 concentration (P) for each of the neighboring places S_N_i (i = 1, 2, …, a), where a is the number of neighboring places. Let us consider that the direction of a neighboring place S_N_i with respect to S_L is ξ. If at the timestamp t the wind is flowing from S_N_i to S_L with the angle of (ξ±11:25)° and the wind speed WS(m/s) is greater than the threshold value ϕ, then the PM2.5 concentration of Y ($P_{i,t+tm}$) for the timestamp (t + tm) is calculated as follows:

$$tm = \lfloor \frac{1}{3600} \left(\frac{L}{S\_N_i(WS_t)} \right) \rfloor \quad (9)$$

$$Y(P_{i,t+tm}) = S\_N_i(P_t) \quad (10)$$

Here, tm is the time required to reach the PM2.5 concentration from S_N_i to S_L, L represents the distance between S_N_i and S_L. $S\_N_i(WS_t)$ and $S\_N_i(P_t)$ represent the wind speed and PM2.5 concentration of S_N_i respectively. The threshold value of the wind speed ϕ is dependent on the distance (L) between the S_L and S_N, and the time required to reach the PM2.5 concentration from S_N to S_L. Here to get the final threshold value ϕ, the NPI algorithm is simulated with various threshold values ranging from WS_{low} to WS_{high}, where WS_{low} and WS_{high} are the lowest and the highest wind speed (m/s) of S_N respectively. The threshold value for which the DLAQPF produces the best prediction result is considered as the final threshold value ϕ of the proposed NPI algorithm.

Now, all these newly generated time-series datasets Y_i(i = 1, 2, … , a) from all the S_Ns are combined with the dataset of S_L based on the matched timestamp to form the input dataset. In the rest of the experiments, this newly generated input dataset is considered.

Symbol	**Meaning**
ξ	Direction of S_N with respect to S_L
ϕ	Threshold wind speed
L	The distance between S_N and S_L
T	Timestamp of each data item of S_N
P	Concentration of PM2.5
WD	Wind direction
WS	Wind speed

Table 2: Notation used in Algorithm 2.

Algorithm 2: Neighboring place's impact (NPI) algorithm

Input: Multivariate time-series S_N {T, P, WD, WS}, ξ and L
Output: Multivariate time-series Y {T, P}

```
for t =1 to M do
    Yt,1 ← S_N ( Tt)
    Yt,2 ← 0
end for
t ← 1
while t < M do
        if (S_N (WD) ≥ ξ – 11.25 and S_N (WD) ≤ ξ +11.25 then
          if (S_N (WS) > ϕ ) then
                  tm ← INT (L/(S_N(WS)/3600))
                  Y t+tm,2 ← S_N(Pt)
          end if
```

```
        end if
        t ← t + 1
end while
```

3.6 Hybrid Prediction Model

The final input dataset of the model consists of meteorological data such as temperature (TEMP), humidity (HUMID), wind direction (WDR), rainfall (RAIN), and, wind speed (WSP), and the AQI of the local station S_L. It also includes the influence of S_Ns' PM2.5 concentrations. The detailed architecture of the proposed hybrid prediction model is presented in Figure 6. The model consists of two major blocks, the first one is CNN (Convolution Neural Network) block and the second one is a GRU(Gated Recurrent Network) block. Finally, the output is generated from the GRU block. The two major blocks of the model are described below:

3.6.1 CNN Block

CNN architectures are excellent for their feature extraction capabilities from the input patches. For that reason, they have produced remarkable success in image processing [Lee and Kwon, 2017]. The CNN can also be enforced in time-series forecasting problems. The recent research [Yang et al., 2015] also proved that it exhibited good performance in analyzing the time-series data. One of the variants of the traditional CNN, one dimensional CNN (1D-CNN) is deployed for air quality time-series forecasting problems.

In the proposed hybrid forecasting model, the CNN block consists of three consecutive layers of 1D-CNN layers, a dropout layer, and a flattened layer. Each 1D-CNN layer comprises convolution sub-layer, activation sub-layer, and pooling sub-layer. Each convolution sub-layer uses a convolution window to process the meteorological and pollutant time-series data to learn the sequence chunk within each input window. In this way, the convolution sub-layer extracts the local trend features from the input multivariate time-series dataset. Then, in the activation sub-layer, the ReLU activation function is applied to the features matrix to increase the nonlinearity of the output. After that, the pooling sub-layer is used. In this layer, the MaxPooling function is used to subsample the input features matrix by selecting the maximum values from the input features matrix. The functions performed by each CNN layer are expressed as follows:

$$\boldsymbol{F_j^p} = \sum \boldsymbol{X_j^{p-1}} \oplus \boldsymbol{P_{j,k}} + \boldsymbol{f W_k^p} \tag{11}$$

$$\boldsymbol{Y_j^p} = \boldsymbol{ReLU}(\boldsymbol{F_j^p}) \tag{12}$$

$$\boldsymbol{X_i^p} = \boldsymbol{Pool}(\boldsymbol{Y_j^p}) \ , \tag{13}$$

where X= the input, P=Filter, W=Bias, p=Number of layers, ⊕= Convolution operation.

Here three 1D-CNN layers are deployed to learn the local trend features. Each layer takes the output from the previous layer and learns the non-linear representation. Then it passes these learned features to the next layer. After the three consecutive 1D-CNN layers, there is a dropout layer. This dropout layer is introduced to remove some of the randomly selected parameters to overcome the overfitting problem. Then the fully connected layer is used to convert the higher-order feature matrix into a feature vector. That feature vector is fed as the input to the GRU block. So, the CNN block not only captures the trend features from the multivariate time-series dataset but also assimilates the spatial correlational features.

Besides the spatial and local trends features extraction capabilities of 1D-CNN, its weight-sharing features reduce the number of learnable parameters for processing the multivariate time-series dataset. And that helps to improve the learning efficiency of the model. Thus, the CNN block can learn more rooted features from the air quality time-series dataset.

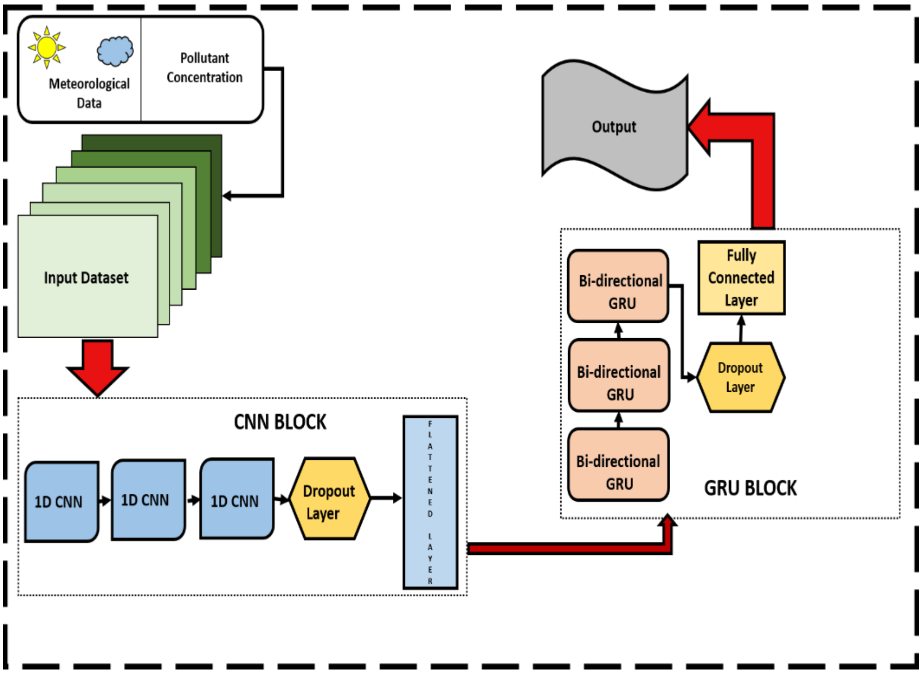


Figure 6: Detailed description of the hybrid prediction model.

3.6.2 GRU Block

In this GRU block, three layers of bi-directional gated recurrent unit (Bi-GRU) followed by a dropout layer and a fully connected layer are used for the prediction of AQI. It is a known fact that the recurrent neural network (RNN) is specially developed to handle sequential data. But the RNN suffers from the vanishing gradient and exploding gradient problem, and due to these problems, it cannot learn the long-term dependencies that exist in the time-series dataset. To solve these problems, two variants of RNN have been developed. These variants are LSTM and GRU. The LSTM can learn the long-term dependencies of the time-series data through its memory cell which comprises three gates (input gate, forget gate, and output gate) [Hochreiter and Schmidhuber, 1997]. The GRU can also learn the long-term dependencies [Bahdanau et al., 2014]. Unlike LSTM, it does not have any memory cell, instead, it has two gates (update gate and reset gate). Due to this simple structure, the GRU architecture has fewer trainable parameters, and its training process is faster compared to the LSTM architecture. Recent research work [Shewalkar, 2019] also proved that the performance GRU is commensurate with LSTM. The operation performed by the GRU unit is expressed as follows:

$$g_t = \omega(K_g * [\beta(t), \alpha(t-1)]) \tag{14}$$

$$rs_t = \omega(K_{rs} * [\beta(t), \alpha(t-1)]) \tag{15}$$

$$\acute{\alpha}(t) = \omega(K_\alpha * [\beta(t), rs_t * \alpha(t-1)]) \tag{16}$$

$$\alpha(t) = (1 - g_t) * \alpha(t-1) + g_t * \acute{\alpha}(t), \tag{17}$$

where g_t is representing the update gate,
ω denotes the activation function,
$\alpha(t)$ and $\alpha(t\text{-}1)$ are representing the input and the previous output respectively,
rs_t represents the reset gate, and
K_g, K_{rs}, K_α are the denoting the weights of the update gate, reset gate and candidate output respectively.

The air quality time-series data can have both forward and backward long-term dependencies. But the ordinary GRU can only learn the long-term forward dependencies. The Bi-GRU is a variant of standard GRU architecture. It can preserve both the long-term dependencies of the time-series data. This Bi-GRU consists of two ordinary GRUs, where the first GRU processes the time-series in the forward direction (t = 1 to T), and the other GRU processes the time-series in the backward direction (t = T to 1). In this way, the Bi-GRUs can extract more useful information that helps to improve the prediction performance of the model.

Here in this GRU block, a series of three Bi-GRU layers are used. Pipelining of Bi-GRU exhibits high efficiency [Li et al., 2018, Lynn et al., 2019]. One Bi-GRU in the GRU block receives the output of another Bi-GRU. In this way, three Bi-GRUs are connected. This kind of alignment helps in extracting higher-order features of the temporal dataset. The pipelining of Bi-GRU delivers the output to the next layer i.e. a dropout layer. The dropout layer randomly selects and discards some features that ultimately overcome the overfitting problem. The next layer in the GRU block is a fully connected layer, which produces the final output. The AQI of the S_L is the final output of this model.

4 Experiment

All the experiments related to this proposed framework are done on a PC with Intel(R) Core (TM) i3-8130U CPU 2.20 GHz, having a 64-bit operating system x64-based processor and memory of 4.00 GB. To perform all the experiments the Python programming language is used. Keras, the open-source deep learning library, has been to make the foundation of a hybrid prediction model.

4.1 Experimental Setup

This work aims to forecast the AQI of S_L one hour ahead by using the meteorological factors and pollutant concentration of the S_Ns. To be specific, here, the previous 24-hours' data are used to predict the AQI of the 25th hour. To carry out this experiment a total of 26,280 data are used. Within the dataset 80% (21,024 data samples) is used for training purposes and the remaining 20% of the data (5,256 data samples) is used for testing purposes. Data normalization is an important step to be followed for a successful and flawless experiment. Data normalization is important as it discards the unwanted repetition of data and to some extent it removes different types of anomalies from the data. The time-series data is highly dynamic and it varies within a wide range. Due to this, the overall learning process slows down. To speed up the learning process and to scale the data within 0 and 1, the data normalization process is carried out. Here, the Min-Max normalization method is used. This normalization technique linearly transforms the data. In this, the minimum and the maximum values from the dataset are picked up and are replaced by using the formula as stated:

$$\boldsymbol{n_{norm}} = \frac{((\boldsymbol{high-low})*(\boldsymbol{n-MinN}))}{(\boldsymbol{MaxN-MinN})}, \quad (18)$$

where MinN and MaxN are the minimum and maximum values of the attribute N in the input dataset.

By using Equation18, the input value n (an attribute of N) is converted to n_{norm}. In the hybrid prediction model, the model parameters are tuned to get the best result for the model. Table 3 represents the parameters used for the proposed hybrid forecasting model.

CNN Block				**GRU Block**				**Epoch**
Filter size			**Dropout**	**No. of neurons**			**Dropout**	
Layer 1	**Layer 2**	**Layer 3**		**Layer 1**	**Layer 2**	**Layer 3**		
72	72	72	0.6	16	16	16	0.6	5-150
72	72	72	0.6	32	32	32	0.6	5-150
72	72	72	0.6	64	64	64	0.6	5-150
72	72	72	0.6	80	80	80	0.6	5-150
72	72	72	0.6	128	128	128	0.6	5-150

Table 3: Parameter settings for the proposed hybrid prediction model.

4.2 Evaluation

In this section, both the imputation algorithm for replacing the missing values in the dataset and the proposed prediction model are assessed to explain their applicability.

4.2.1 Evaluation of the proposed imputation algorithm for missing value replacement

There are several imputation algorithms already existing. They are widely used in replacing the missing values in the datasets. Some of the well-known imputation algorithms are the Mean/Mode Imputation algorithm [Tsai et al., 2018], autoregressive imputation algorithm [Bashir and Wei, 2018], maximum likelihood imputation algorithm [Enders, 2001], K-Nearest Neighbor imputation algorithm [Batista et al., 2002] and bagging algorithm [Andiojaya and Demirhan, 2019]. To examine the applicability of the proposed imputation algorithm, it is compared with the existing imputation algorithms. The results are shown in Table 4.

4.2.2 Evaluation of the prediction model

After the construction of any model, it is very important to evaluate its performance. Any type of regression model must also go through the evaluation process through which the errors of the model can be filtered and a comparative study of the performance of the proposed model with the established models can be achieved. For evaluating the efficiency of the hybrid prediction model here three evaluation metrics are used. These are as follows:

- **MAE (Mean Absolute Error):** This error evaluating technique is too popularly used to foretell the errors within a time-series analysis. During the training, the process to determine the loss function MAE is used. The MAE can better showcase the actual scenario of the errors committed during the training process of the prediction framework. MAE is calculated as,

$$MAE = \frac{1}{N}\sum_{i=1}^{N} |X_i - Y_i| \tag{19}$$

- **RMSE (Root Mean Square Error)**: It helps in observing the difference between the predicted value and the perceived value. A good prediction model will always exhibit a smaller RMSE value. So, RSME is calculated as,

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(X_i - Y_i)^2} \tag{20}$$

- **SMAPE (Symmetric Mean Absolute Percentage Error)**: This error evaluation method evaluates a model by considering percentage errors. So, SMAPE is defined using the following formula:

$$SMAPE = \frac{100\%}{N}\sum_{i=1}^{N}\frac{|X_i - Y_i|}{(|X_i| + |Y_i|)/2} \quad , \tag{21}$$

where X_i=the predicted value, Y_i=the actual value, i= every point of observation, N= the total number of observations. To determine the effectiveness of our proposed work, the same framework is compared with some standard models like:

- SVR (Support Vector Regressor)
- Stacked LSTM: Three Bi-Directional LSTM layers are used to analyze both the forward and backward time-series data.
- GRU: Three layers of Bi-GRU are used.

State of the art models like

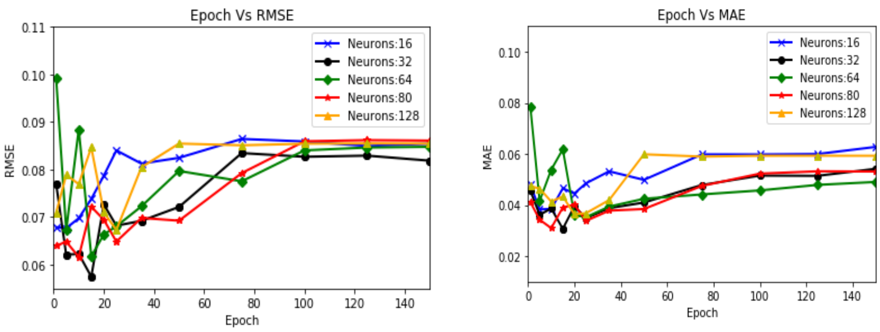
- CBGRU [Tao et al., 2019]
- DAQFF [Du et al., 2021]

5 Results & Discussion

5.1 Simulation of DLAQPF

To get the optimal result of the proposed DLAQPF framework, DLAQPF was simulated based on the parameter settings presented in Table 3. Figure 7 displays the performance graph of the proposed framework DLAQPF with the different number of neurons and a varied number of epochs. From there, it is clearly observed that DLAQPF is delivering more accurate prediction results with 32 numbers of neurons and if the number of the epoch is 16. Here to overcome the overfitting problem, a dropout rate of 0.6 is used for both the dropout layers of CNN and GRU blocks. For the rest of the experiments, these parameter settings are used.

To get the optimum threshold value of wind speed (ϕ) for each of the neighboring stations S_N_i (i = 1, 2, 3, …, a; where a is the number of neighboring stations) the DLAQPF is simulated with varying threshold values of wind speeds for the Neighboring Place's Impact (NPI) algorithm. From the experiments the threshold value ϕ for the neighboring station Richmond is set to 1.74 m/s, for St. Marry it is set to 2.05 m/s, for Bringelly it is initialized to 1.77 m/s, for Liverpool it is finalized to 1.93 m/s, and for Randwick it is assigned to 1.76 m/s.



(a) RMSE.

(b) MAE.

Figure 7: Performance evaluation (a) RMSE (b) MAE.

5.2 Effectiveness of proposed imputation algorithm for replacing missing values

To ascertain the effectiveness of the proposed imputation algorithm, randomly 5%, 10%, 15%, 20%, and 30% data are removed from the multivariate time-series dataset, and to replace those missing values, the well-established imputation algorithms are used. Depending on the results produced by the prediction model, the RMSE and MAE are calculated. Table 4, depicts the performance of various imputation algorithms on the removal of those aforesaid portions of data from the dataset. From this table, it can be observed that as the percentage of missing values increases, errors of all the imputation algorithms increase. Further, its increase in prediction error rates is much lower than that of the other imputation algorithms with an increasing percentage of missing values. Whereas the Mean/Mode imputation algorithm produces the worst performance and its rate of increase in error rates is higher with the increasing percentage of missing values. It can be seen that the proposed Seasonality-based imputation algorithm produces the lowest errors (RMSE and MAE) for all the percentages of missing values compared to other algorithms except the Bagging algorithm. When the missing value percentage is greater than or equal to 15% then the Bagging algorithm is performing slightly better than the proposed Seasonality-based imputation algorithm. But it can further be observed that when the missing value is greater or equal to 15%, its RMSE and MAE are slightly more than that of only the Bagging algorithm. From this discussion it can be concluded that the Seasonality-based imputation algorithm proposed in this letter is also acceptable for missing values more than 10%. The dataset used in this proposed work is having lesse than 10% missing values. So the proposed seasonality based imputation algorithm is best suited and in the entire experiment the proposed Seasonality-based imputation algorithm is used. The reason behind the better performance of the proposed imputation algorithm is that it learns the seasonality and handles the missing values accordingly.

Algorithm	**% of missing values**									
	5%		**10%**		**15%**		**20%**		**30%**	
	RMSE	**MAE**	**RMSE**	**MAE**	**RMSE**	**MAE**	**RMSE**	**MAE**	**RMSE**	**MAE**
Mean/ Mode	3.8215	3.0851	4.6113	4.9502	5.9950	6.2839	6.7809	7.2408	7.3369	8.3321
Auto regressive	2.8112	2.4508	3.3619	3.0146	3.8041	3.9876	4.9112	5.1256	5.7445	6.2141
Maximum likelihood	3.1471	2.9613	3.4206	3.4876	4.0019	4.7204	5.4811	6.8872	6.2553	7.1145
K-NN	3.2048	2.8440	3.3966	3.5016	4.0283	4.9214	5.5296	6.7821	6.3225	7.5258
Bagging algorithm	2.8058	2.3159	3.1967	2.9430	3.6914	3.8049	4.8204	4.9571	5.3159	6.2214
Proposed algorithm	**2.7061**	**2.2593**	**2.8493**	**2.8245**	**3.7628**	**3.9612**	**4.9614**	**5.0085**	**5.4213**	**6.3281**

Table 4 : Performance evaluation of proposed imputation algorithm.

5.3 Performance evaluation considering the air quality of the neighboring places

Experiments are carried out in two different phases to examine whether there are influences of the air quality of the neighboring places in determining the air health of a particular place. The two phases are:

- AQI of S_L is predicted by considering the pollutant concentration of S_Ns.
- AQI of S_L is predicted without considering the pollutant concentration of S_Ns.

Table 5 shows that when the pollutant concentrations of the S_Ns are considered, the experiment shows a better result in predicting the AQI of S_L. On the other hand, it is clear that the error rates are higher when the influence of the pollutant concentrations of S_Ns is ignored in the diagnosis of air health in S_L.

Metric	RMSE	MAE	SMAPE
Considering the influence of the neighboring places	2.7013	2.2730	8.9340
Ignoring the influence of the neighboring places	3.4621	2.8932	10.8762

Table 5: Performance comparison with and without considering effects of neighboring places.

5.4 Comparative study with other models

To evaluate the accuracy of the results delivered by our proposed model DLAQPF, the RMSE, MAE, and SMAPE are calculated for all the existing models like SVR, Bi-LSTM, Bi-GRU, CBGRU, and DAQFF. Table 6 shows the performance evaluation of the proposed model and Bi-LSTM, SVR, Bi-GRU, CB-GRU, and DAQFF. In Table 6, it can be seen that the proposed model DLAQPF is committing the least error among all the models.

To be more critical about the prediction models, a comparison with the same dataset of New South Wales, Australia is carried out and those performance graphs are exhibited in Figure 8. There, the predicted AQI versus actual AQI is plotted. The same dataset is applied to the shallow models like LSTM, SVR, and GRU and the two-state of the art models like CBGRU and DAQFF. Here, the result is zoomed in for the 1000 data points. The data point is starting from 02/07/2019 01:00 hours to 12/08/2019 16:00 hours is shown for each model. From the plotted graphs of the three traditional shallow forecasting models Figures 8a, 8b, and 8c it can be observed that shallow deep learning models LSTM and GRU are performing better than the traditional machine learning model SVR, as, from the three graphs, it is clear that in LSTM and GRU wave peaks and wave valleys are agreeing more often than in SVR. So, it can be concluded that the shallow deep learning models like LSTM and GRU are better than the shallow machine learning model SVR.

Now, if Table 6 is consulted, it can be seen that GRU always has the lowest error in comparison with LSTM and SVR. Comparison among the shallow deep learning architecture and the combination of more than one deep learning architecture can be showcased in Figure 8. It is clear from Figure 8 that CBGRU (which is a combination of 1D-CNN and Bi-GRU), and DAQFF (which is a combination of LSTM and 1D-CNN) show better prediction accuracy than the deep learning architecture GRU. If Figures 8a, 8b, 8c, 8d, 8e, and 8f are observed, then it can be clearly concluded that the proposed model DLAQPF is outperforming the other models. For both the actual and predicted values the wave peaks and the wave valleys are agreeing better than any other model depicted. To display the performance of the various forecasting models, the boxplot deviation analysis is shown in Figure 9. The difference between the actual result and the predicted result is the measured deviation. The variation of the deviating data is represented by the height of the box. To represent more centralized data, fatter boxes and shorter whiskers are used. The plotting clearly shows the proposed model DLAQPF and CBGRU have notches near 0 which means their medians are nearby 0. If concentrated on the boxplot, it can easily be seen that among all the forecasting models plotted in Figure 9, the DLAQPF has the fattest box with the shortest whiskers. Thus, from this observation, it can be concluded that DLAQPF delivers the highest accuracy in forecasting the AQI of a place.

The key factor which is responsible for the success of the proposed model is the combination of more than one DL architecture in predicting the AQI. DLAQPF is outperforming all other models for two basic reasons:

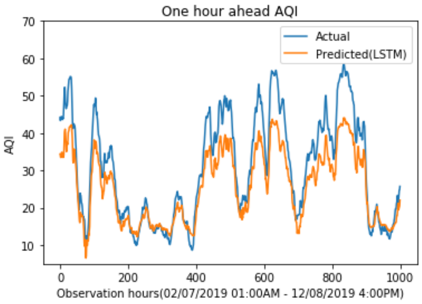
- DLAQPF can address not only the pollution concentration and meteorological factors of S_L but it addresses the PM2.5 and other meteorological factors of the nearby S_Ns to predict the AQI of S_L.
- As a building block, the series of CNN layers allows the model to extract more important local complex features from the input window. Moreover, the GRU block is used in DLAQPF. The GRU block comprises a stack of three Bi- GRU, which is capable of abstracting the temporal feature as well as the forward and backward (both ways) dependencies of the time-series data.

Model	**RMSE**	**MAE**	**SMAPE**
LSTM	6.1163	5.2208	18.1288
SVR	6.8594	5.6024	17.7777
GRU	4.7207	3.8146	12.6980
CBGRU	3.4459	2.8652	10.8251
DAQFF	3.9193	3.2155	13.7465
DLAQPF	2.7013	2.2730	08.9340

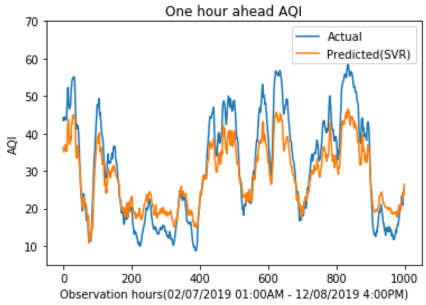
Table 6 : Performance evaluation of different models.

6 Conclusion & Future Scope

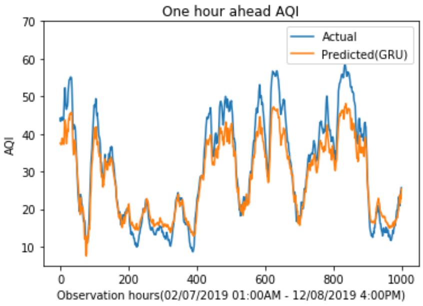
In this work, the focus was to develop a forecasting model using meteorological data which would deliver accurate and stable prediction results. The work was to forecast the AQI of a particular place S_L by considering both the pollutant concentration and meteorological factors of local and the pollutant concentration of the neighboring places denoted as S_Ns. The main contributions of this paper are:



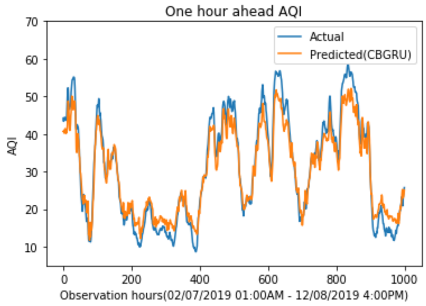
(a) LSTM.



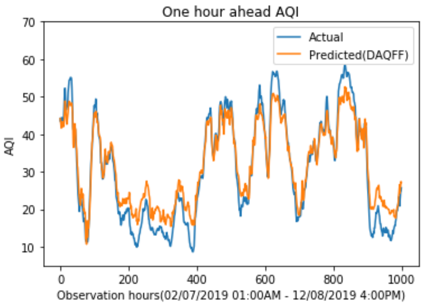
(b) SVR.



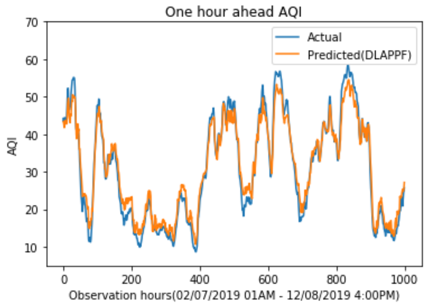
(c) GRU.



(d) CBGRU.



(e) DAQFF



(f) DLAQPF

Figure 8: Performance graph (a) LSTM (b) SVR (c) GRU (d) CBGRU (e) DAQFF (f) DLAQPF.

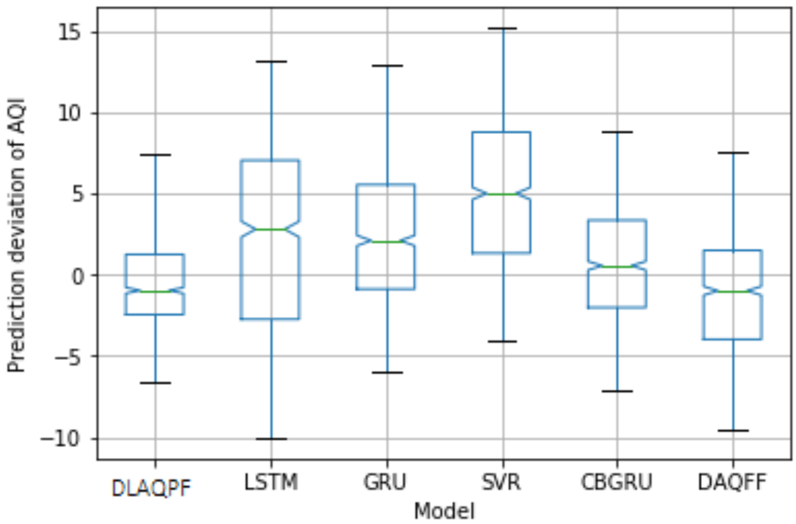


Figure 9: The boxplot to depict the prediction deviation of different forecasting models.

1. Firstly, in this work, a framework for building a forecasting model for a place with the ability to address not only the pollutant concentration but also the meteorological factors of itself and the nearby places was proposed.
2. Secondly, a seasonality-based imputation algorithm is proposed in this letter to impute the missing values for generating a more convenient dataset.
3. The proposed model DLAQPF was built in a hybrid fashion by deploying stacked 1D-CNN and stacked Bi-GRU. The model is exploring the advantages of both the building blocks. The 1D-CNN enables expert feature extraction and it's good in dimensionality reduction, whereas, the Bi-GRU is an expert in abstracting temporal features and it addresses the both-way dependencies of the time-series data.
4. After several experimentations and comparisons, it was proved that the proposed model has better prediction capability while showing the least error rates. Moreover, the work also confirms that the AQI of a particular place gets influenced by the pollutant concentration of the neighboring places.

Thus this research work validates the two research questions (RQ1 and RQ2) raised in Section 1.

As the prediction result exhibited by this proposed work is more accurate in comparison with other prediction models, the prediction result may be used by the citizens of the smart city to be cautious. According to the forecast, they may plan for some alternatives while going outdoors in such a way that they would not expose themselves to the polluted area. From the broader perspective, the prediction result may be used by the governing authorities like municipalities to take some precautionary steps to warn their fellow citizens so that the citizens may escape from the worst effects of pollution. The warnings may be in the form of some early-alert generation or in the form of some actions like traffic diversion, new route plan proposals, and similar kinds of proposals. Thus, it can be expected that the outcome of this work will be helpful to the people residing in smart cities as well as to the policymakers. The proposed model is dependent on historical data. If this model is to be used for predicting the AQI of a place that lacks historical data, then it is impossible to forecast the AQI of that place. So, in future research, the transfer learning technology may be incorporated to build the historical dataset so that the model can perform as per expectation.

References

[Ameer et al., 2019] Ameer, S., Shah, M. A., Khan, A., Song, H., Maple, C., Islam, S. U., Asghar, M. N.: 'Comparative analysis of machine learning techniques for predicting air quality in smart cities'; IEEE Access, 7 (2019), 128325–128338.

[Andiojaya and Demirhan, 2019] Andiojaya, A., Demirhan, H.: 'A bagging algorithm for the imputation of missing values in time series'; Expert Systems with Applications, 129 (2019), 10–26.

[AustraliaNewSouthWales, 2022] AustraliaNewSouthWales: 'Australia'; (2022).

[Bahdanau et al., 2014] Bahdanau, D., Cho, K., Bengio, Y.: 'Neural machine translation by jointly learning to align and translate'; ArXiv Preprint ArXiv:1409.0473 (2014).

[Bashir and Wei, 2018] Bashir, F., Wei, H.-L.: 'Handling missing data in multivariate time series using a vector autoregressive model-imputation (VAR-IM) algorithm'; Neurocomputing, 276 (2018), 23–30.

[Batista et al., 2002] Batista, G. E., Monard, M. C., others: 'A study of K-nearest neighbour as an imputation method.'; His, 87, 251–260 (2002), 48.

[Bekkar et al., 2021] Bekkar, A., Hssina, B., Douzi, S., Douzi, K.: 'Air-pollution prediction in smart city, deep learning approach'; Journal of Big Data, 8, 1 (2021), 1–21.

[Che et al., 2020] Che, W., Frey, H. C., Fung, J. C. H., Ning, Z., Qu, H., Lo, H. K., et al.: 'PRAISE-HK: A personalized real-time air quality informatics system for citizen participation in exposure and health risk management'; Sustainable Cities and Society, 54 (2020), 101986.

[Deleawe et al., 2010] Deleawe, S., Kusznir, J., Lamb, B., Cook, D. J.: 'Predicting air quality in smart environments'; Journal of Ambient Intelligence and Smart Environments, 2, 2 (2010), 145–154.

[Direction, 2022] Direction: 'Directions'; (2022).

[Donnelly et al., 2015] Donnelly, A., Misstear, B., Broderick, B.: 'Real time air quality forecasting using integrated parametric and non-parametric regression techniques'; Atmospheric Environment, 103 (2015), 53–65.

[Du et al., 2021] Du, S., Li, T., Yang, Y., Horng, S. J.: 'Deep Air Quality Forecasting Using Hybrid Deep Learning Framework'; IEEE Transactions on Knowledge and Data Engineering, 33, 6 (2021), 2412–2424. https://doi.org/10.1109/TKDE.2019.2954510

[Enders, 2001] Enders, C. K.: 'A primer on maximum likelihood algorithms available for use with missing data'; Structural Equation Modeling, 8, 1 (2001), 128–141.

[Gamboa, 2017] Gamboa, J. C. B.: 'Deep learning for time-series analysis'; ArXiv Preprint ArXiv:1701.01887 (2017).

[Gao et al., 2019] Gao, X., Li, X., Zhao, B., Ji, W., Jing, X., He, Y.: 'Short-term electricity load forecasting model based on EMD-GRU with feature selection'; Energies, 12, 6 (2019), 1140.

[Ghose and Rehena, 2020] Ghose, B., Rehena, Z.: 'A mechanism for air health monitoring in smart city using context aware computing'; Procedia Computer Science, 171 (2020), 2512–2521.

[Ghose and Rehena, 2021] Ghose, B., Rehena, Z.: 'A Deep Learning Approach for Predicting Air Pollution in Smart Cities'; In Computational Intelligence and Machine Learning. Springer (2021), 29–38.

[Hochreiter and Schmidhuber, 1997] Hochreiter, S., Schmidhuber, J.: 'Long short-term memory'; Neural Computation, 9, 8 (1997), 1735–1780.

[Honarvar and Sami, 2019] Honarvar, A. R., Sami, A.: 'Towards sustainable smart city by particulate matter prediction using urban big data, excluding expensive air pollution infrastructures'; Big Data Research, 17 (2019), 56–65.

[Kim et al., 2015] Kim, K.-H., Kabir, E., Kabir, S.: 'A review on the human health impact of airborne particulate matter'; Environment International, 74 (2015), 136–143.

[Kurt et al., 2008] Kurt, A., Gulbagci, B., Karaca, F., Alagha, O.: 'An online air pollution forecasting system using neural networks'; Environment International, 34, 5 (2008), 592–598.

[Lee et al., 2015] Lee, A., Szpiro, A., Kim, S. Y., Sheppard, L.: 'Impact of preferential sampling on exposure prediction and health effect inference in the context of air pollution epidemiology'; Environmetrics, 26, 4 (2015), 255–267.

[Lee and Kwon, 2017] Lee, H., Kwon, H.: 'Going deeper with contextual CNN for hyperspectral image classification'; IEEE Transactions on Image Processing, 26, 10 (2017), 4843–4855.

[Leng et al., 2020] Leng, J., Wang, Q., Liu, K.: 'Sustainable design of courtyard environment: From the perspectives of airborne diseases control and human health'; Sustainable Cities and Society, 62 (2020), 102405.

[Li et al., 2018] Li, L., Cai, G., Chen, N.: 'A rumor events detection method based on deep bidirectional GRU neural network'; In 2018 IEEE 3rd International Conference on Image, Vision and Computing (ICIVC) (2018), 755–759.

[Liu et al., 2021a] Liu, B., Yu, X., Chen, J., Wang, Q.: 'Air pollution concentration forecasting based on wavelet transform and combined weighting forecasting model'; Atmospheric Pollution Research, 12, 8 (2021a), 101144.

[Liu et al., 2021b] Liu, D.-R., Hsu, Y.-K., Chen, H.-Y., Jau, H.-J.: 'Air pollution prediction based on factory-aware attentional LSTM neural network'; Computing, 103, 1 (2021b), 75–98.

[Liu et al., 2019] Liu, H., Li, Q., Yu, D., Gu, Y.: 'Air quality index and air pollutant concentration prediction based on machine learning algorithms'; Applied Sciences, 9, 19 (2019), 4069.

[Lynn et al., 2019] Lynn, H. M., Pan, S. B., Kim, P.: 'A deep bidirectional GRU network model for biometric electrocardiogram classification based on recurrent neural networks'; IEEE Access, 7 (2019), 145395–145405.

[Martínez-España et al., 2018] Martínez-España, R., Bueno-Crespo, A., Timon-Perez, I. M., Soto, J. A., Ortega, A. M., Cecilia, J. M.: 'Air-Pollution Prediction in Smart Cities through Machine Learning Methods: A Case of Study in Murcia, Spain.'; J. Univers. Comput. Sci., 24, 3 (2018), 261–276.

[Minoli et al., 2017] Minoli, D., Sohraby, K., Occhiogrosso, B.: 'IoT considerations, requirements, and architectures for smart buildings—Energy optimization and next-generation building management systems'; IEEE Internet of Things Journal, 4, 1 (2017), 269–283.

[NSW, 2022] NSW: 'Dataset: New South Wales'; (2022).

[Qianrao, 2016] Qianrao, F.: 'Research on Haze Prediction Based on Multivariate Linear Regression [J]'; Computer Science, 43, S1 (2016), 526–528.

[Rehena and Janssen, 2019] Rehena, Z., Janssen, M.: 'The smart city of Pune'; In Smart City Emergence. Elsevier (2019), 261–282.

[Ren et al., 2021] Ren, X., Li, X., Ren, K., Song, J., Xu, Z., Deng, K., Wang, X.: 'Deep learning-based weather prediction: a survey'; Big Data Research, 23 (2021), 100178.

[Rouse, 2019] Rouse, M.: 'What is a Smart City? Definition from WhatIs.com.'; (2019, July).

[Schürholz et al., 2020] Schürholz, D., Kubler, S., Zaslavsky, A.: 'Artificial intelligence-enabled context-aware air quality prediction for smart cities'; Journal of Cleaner Production, 271 (2020), 121941.

[Shewalkar, 2019] Shewalkar, A.: 'Performance evaluation of deep neural networks applied to speech recognition: RNN, LSTM and GRU'; Journal of Artificial Intelligence and Soft Computing Research, 9, 4 (2019), 235–245.

[Tao et al., 2019] Tao, Q., Liu, F., Li, Y., Sidorov, D.: 'Air pollution forecasting using a deep learning model based on 1D convnets and bidirectional GRU'; IEEE Access, 7 (2019), 76690–76698.

[Tsai et al., 2018] Tsai, C.-F., Li, M.-L., Lin, W.-C.: 'A class center based approach for missing value imputation'; Knowledge-Based Systems, 151 (2018), 124–135.

[USEPA, 2020] USEPA: 'US-EPA'; (2020).

[Vardoulakis et al., 2003] Vardoulakis, S., Fisher, B. E. A., Pericleous, K., Gonzalez-Flesca, N.: 'Modelling air quality in street canyons: a review'; Atmospheric Environment, 37, 2 (2003), 155–182.

[Verma et al., 2018] Verma, I., Ahuja, R., Meisheri, H., Dey, L.: 'Air pollutant severity prediction using Bi-directional LSTM Network'; In 2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI) (2018), 651–654.

[Wang et al., 2019] Wang, B., Kong, W., Guan, H., Xiong, N. N.: 'Air quality forecasting based on gated recurrent long short term memory model in Internet of Things'; IEEE Access, 7 (2019), 69524–69534.

[Wei et al., 2015] Wei, Y., Xia, W., Lin, M., Huang, J., Ni, B., Dong, J., et al.: 'HCP: A flexible CNN framework for multi-label image classification'; IEEE Transactions on Pattern Analysis and Machine Intelligence, 38, 9 (2015), 1901–1907.

[Worldmap, 2022] Worldmap: 'World map'; (2022).

[Wu and Lin, 2019] Wu, Q., Lin, H.: 'Daily urban air quality index forecasting based on variational mode decomposition, sample entropy and LSTM neural network'; Sustainable Cities and Society, 50 (2019), 101657.

[Yang et al., 2015] Yang, J., Nguyen, M. N., San, P. P., Li, X. L., Krishnaswamy, S.: 'Deep convolutional neural networks on multichannel time series for human activity recognition'; In Twenty-fourth international joint conference on artificial intelligence (2015).

[Yang et al., 2016] Yang, W., Lin, L., Tian, L.: 'The haze weather analysis and prediction based on the wavelet analysis'; J. Shaanxi Univ. Sci. Technol.(Natural Sci. Ed.), 34, 3 (2016), 166–170.

[Yi et al., 2018] Yi, X., Zhang, J., Wang, Z., Li, T., Zheng, Y.: 'Deep distributed fusion network for air quality prediction'; In Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery \& data mining (2018), 965–973.

[Zhang et al., 2012] Zhang, Y., Bocquet, M., Mallet, V., Seigneur, C., Baklanov, A.: 'Real-time air quality forecasting, part I: History, techniques, and current status'; Atmospheric Environment, 60 (2012), 632–655.

[Zhu et al., 2017] Zhu, S., Lian, X., Liu, H., Hu, J., Wang, Y., Che, J.: 'Daily air quality index forecasting with hybrid models: A case in China'; Environmental Pollution, 231 (2017), 1232–1244.