Urban Climate 36 (2021) 100800

Contents lists available at ScienceDirect

Urban Climate

journal homepage: www.elsevier.com/locate/uclim

Multi-directional temporal convolutional artificial neural network for PM2.5 forecasting with missing values: A deep learning approach

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

[a] Department of Computer science and Engineering, National Institute of Technology Rourkela, 769008, India
[b] Department of Electronics and Communication Engineering, National Institute of Technology Rourkela, 769008, India

ARTICLE INFO

Keywords:
Deep learning
Time series forecasting
Temporal Convolutional Network
Artificial Neural Network
PM2.5

ABSTRACT

Data imputation and forecasting are the major research areas in environmental data engineering. Solving those critical issues has an immense impact on air pollution management, consequently improving social, economic growth, and public health. Missing data is a common issue for all the domains, especially for environmental data analysis. Most of the research study tries to solve all these problems of time series data using different models. This research study presents a novel deep learning-based hybrid model architecture to solve these issues in a single training process. We come up with Multi-directional Temporal Convolutional Artificial Neural Network (MTCAN) model to impute and forecast PM2.5 pollutant concentration in a single training process. The main idea of the multi-directional properties of MTCAN is to interpolate the PM2.5 pollutant feature matrix to impute its value. Ultimately, it maintains the temporal correlation within the features' measurement and meteorological and pollutant variables to impute PM2.5 missing values. The MTCAN model performs feature learning and sequential modeling simultaneously with a wide range of past observations for long-term forecasting, minimizing memory size requirement and training cost. Experimental results indicate that the proposed model is superior to baseline pollution forecasting models, which prove its effectiveness in air quality modeling.

1. Introduction

With the rapid acceleration of industrialization and urban development, air pollution has become a serious problem that negatively impacts both the human body and the environment. Therefore, air pollution forecasting is considered a major issue for human health and environmental protection (Perez et al. 2020). Air quality monitoring and modeling have become equally important to ensure that the public gets real-time health alerts and advice based on their areas' pollution level. Accordingly, government and policymakers follow the strategy to control some pollutants on par with world Health Organization Standards (Han et al. 2020). Particulate matter pollutants, including PM2.5 (PM with a diameter of fewer than 2.5 μm) and PM10 (PM with a diameter of fewer than 10 μm), are the major contributors to increasing air pollution, which has caused 4.2 million deaths worldwide (Kök et al. 2017). Industrial sources are the primary sources of PM2.5 and PM10. Particulate matter is the hazardous pollutants among all pollutants as it is easy to inhale and can worsen the respiratory systems of the human body (Pozzer et al. 2019). Pollutant forecasting based on weather conditions is more

* Corresponding author.
E-mail address: 517cs6019@nitrkl.ac.in (K.K.R. Samal).

https://doi.org/10.1016/j.uclim.2021.100800
Received 11 November 2020; Received in revised form 4 February 2021; Accepted 4 February 2021
Available online 13 February 2021
2212-0955/

reliable and accurate (Zhang et al. 2019b). This study's primary goal is to enhance forecasting sensitivity of PM2.5 over long-term duration rather than the short term duration with weather conditions. Gucheng, one of China's polluted areas, and Talcher coal-field, an Indian air pollution monitoring site with their weather conditions, has considered as the study areas in this research study.

Many air pollution forecasting approaches have been proposed, which includes mainly statistical, shallow machine learning, and deep learning methods. These statistical models have limited forecasting accuracy due to their inability to handle multivariate and nonlinear time series dataset (Tao et al. 2019). Air pollution is usually affected by weather, climate variables, traffic, and other factors; it must consider these parameters for air pollution forecasting. Simultaneously, statistical and machine learning models fail to consider those parameter correlations and substantial multivariate air quality datasets (Du et al. 2018).

In recent years, with the rapid growth of artificial intelligence and big data techniques, deep learning has become an active research field for air quality forecasting. Data-driven deep learning techniques are the most famous techniques in computer vision, image processing, video analytic, sequential modeling, and so on. Though these data-driven approaches can extract the past dataset's hidden features and perform sequential modeling more significantly, these models may suffer from noisy and incomplete missing data. Missing data is a common problem for all environmental data analysis (Karmitsa et al. 2020), especially for sensor generated air quality data forecasting. This type of issue mostly occurs due to sensor shutdown, system crashes, environmental disasters, and so on. Imputation is a method for recovering those missing values (Li et al. 2015; Song et al. 2019). Past research studies usually ignore the incomplete data for air quality modeling, which ultimately yields ineffective forecasting results (Zhu et al. 2019). So incorporating the domain knowledge may increase forecasting accuracy. Therefore this research study focuses on both missing data and forecasting issues of air pollution data modeling. Overall, the main contributions of the research study are as follows,

1. We develop effective multi-direction imputation techniques that make use of correlation among the different parameters of pollutant and meteorological factors at different timestamps and consider the correlations among other measurements of each parameter. Ultimately, it exploits the correlation between each parameter's measurement and uses the correlation among each feature. Thus, the correlation between the pollutant and meteorological factors is utilized to impute PM2.5 missing values.
2. The MTCAN model integrates the rapid feature extraction capability of Convolution Neural Network (CNN) and the sequential temporal modeling feature of Recurrent Neural Network (RNN) in parallel to improve the forecasting accuracy. Dilated convolution with residual mapping approach is adopted to consider a wide range of massive past data, so there will be no information loss from the future to the past.
3. Prediction accuracy is further enhanced by using a fine-tuning layer. This layer adjusts the neuron weight of previously trained forecasting model based on the predicted and observed values correlations.

The rest of this research study is organized as follows. In Section 2, we review literature surveys on data imputation techniques and air quality modeling approaches. Section 3 formulates the problem behind this study. In Section 4, we describe the proposed deep learning architecture framework in detail. Section 5 compares the proposed model performance with state-of-the-art models and discuss their results. Finally, Section 6 concludes this research work.

2. Related work

2.1. Data imputation technique

Air quality modeling with the inconsistent missing data is one of the major issues (Quinteros et al. 2019) of environmental modeling. Ultimately, it is a data validation task (Wu et al. 2020b). One method is to impute the missing value by using partially available other parameter value and the second method is to recover the missing attributes by imputed predicted values. Methods for recovering missing values are data interpolation (Yen et al. 2019), data imputation (Yoon et al., 2018). Frequently used techniques are interpolation such as linear (Caillault et al. 2017), spline and cubic (Gnauck 2004). Interpolation techniques usually use mean, median, and mode of partially existing parameter value to predict the missing values; these techniques are important where each parameter value is important. These techniques try to reconstruct the data by capturing temporal correlation within the data stream only but ignore the other streams' correlation. In contrast, imputation techniques capture the data-streams temporal correlation but ignore the measurements' temporal correlations within the data-stream.

Imputation techniques are categorized into two types, i.e., simple imputation techniques such as mean, median, and mode imputation techniques. Second type of imputation technique is multiple imputation technique like maximum likelihood, expectation-maximization (EM) (Bashir and Wei 2018; Rumaling et al. 2020; Mustafa et al. 2011) and KNN based imputation (Dixon 1979; Batista et al. 2002; Malarvizhi and Thanamani 2012; Junninen et al. 2004; Beretta and Santaniello 2016; Yang and Chiang 2020). Single imputation techniques basically utilized for univariate dataset as they do not employ additional parameters to recover the missing values, whereas multiple imputation techniques implemented inter-variables correlations to recover those missing values (Rantou 2017). The past research studies indicate that the meteorological variables and other parameters have direct and indirect significant impacts on pollutant concentration. Hence, multiple imputation techniques can improve air quality data, which ultimately results in better pollution forecasting.

2.2. Air quality modeling

Existing air quality modeling techniques are generally categorized into knowledge-based and data-driven techniques (Xu and

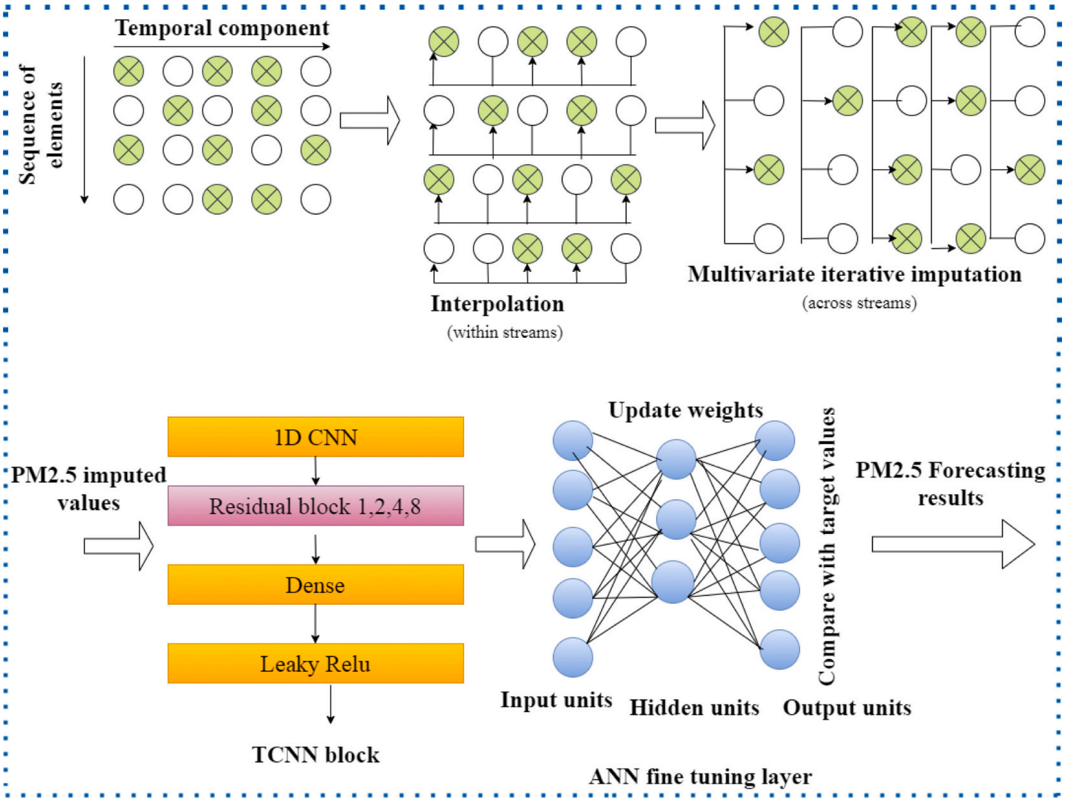

Fig. 1. Porposed MTCAN architecture.

Yoneda 2019). The knowledge-based models implement chemical processes for air quality modeling. Many of the knowledge-based techniques are still using for air quality modeling. However, knowledge-based air quality modeling needs vital information and experience in atmospheric science (Cobourn 2010). These models also have too much computational cost. Past studies in knowledge-based approaches have proven that these models have poor forecasting performance than data-driven models. It has been seen from past research studies that dynamic and high complex strategies followed by knowledge-based models are the main reason for their poor performance. On the other hand, data-driven air pollution modeling can be performed using statistical techniques based on historical datasets, which analyzes air pollution forecasting data's statistical pattern. Statistical prediction models include Autoregressive Integrated Moving Average (ARIMA) (Xu and Yoneda 2019), and Seasonal Autoregressive Integrated Moving Average (SARIMA), linear interpolation, Principle Component Analysis (PCA) techniques, and so on (Han et al. 2020; Samal et al. 2019). On the other hand, shallow machine learning models such as Fbprophet, Artificial Neural Network (ANN), Support Vector Machine (SVM), Support Vector Regression (SVR) (Tao et al. 2019; Murillo-Escobar et al. 2019), and Radial Basis Function (RBF) can be utilized for nonlinear air pollution modeling in a high dimensional space. These shallow conventional machine learning and statistical models ignore pollutants' temporal relationship and other affecting meteorological factors (Qin et al. 2019; Samal et al. 2019). The study of the deep learning approach in multidisciplinary areas has overcome these types of issues. Deep learning models can handle massive datasets and can easily extract complex correlations among the extracted features such as Convolutional Neural Network (CNN) (Du et al., 2018; Qin et al. 2019; Tao et al. 2019), Long Short Term Neural Network (LSTM) (Du et al. 2018; Xu and Yoneda 2019; Qin et al. 2019; Tao et al. 2019), Gated Recurrent Unit (GRU) (Du et al. 2018; Tao et al. 2019), Bidirectional Long Short Term Neural Network (BILSTM) (Du et al. 2018), Bidirectional Gated Recurrent Unit (BIGRU) (Tao et al. 2019) models and so on. These models can handle long sequential datasets by training and adjusting weights, which ultimately reduced the forecasting errors and improved performance. However, these models face difficulties in solving the missing data and forecasting issues with a large meteorological dataset, so we introduced the proposed model to focus on these areas and solve all these issues in a single training process.

3. Problem formulation

we are given an input sequence $x_0, x_1, x_2. \ldots x_T$, and aim to forecast some corresponding outputs $y_0, y_1, y_2. \ldots y_T$ at each time, where x has null values for some t. The main objective is to forecast the output y_T for some time T, we are usually restricted to use only previously observed values as inputs i.e. $x_0, x_1, x_2. \ldots x_t$. Formally, a sequential modeling neural network network is a function $F : X^{T+1} \rightarrow Y^{T+1}$ which produces the mapping $\widetilde{y}_0, \widetilde{y}_1, \widetilde{y}_2 \ldots . \widetilde{y}_T = F(x_0, x_1, x_2 \ldots . x_T)$, if it satisfies the causal constraint that y_T depends only on x_0, $x_1, x_2. \ldots x_t$ and not on any future inputs $(x_{t+1}, x_{t+2}, x_{t+3}. \ldots x_T)$. The main aim of the sequential modeling setting is to find a function F

that reduces some expected loss between the actual outputs and the forecasting outputs, $L(y_0, y_1, y_2. \ldots y_T, F(x_0, x_1, x_2. \ldots x_T))$, where the input sequences and prediction outputs are drawn according to some distribution.

4. Methodology framework

In the following, we discuss the proposed neural network-based long term pollution forecasting model. Our proposed Multi-directional Temporal Convolutional Artificial Neural Network (MTCAN) model's multi-directional properties include both interpolation and imputation block simultaneously instead of only one. It operates in both forward and backward directions. It considers both the temporal correlation within each variable measurement and the variables to impute the target variable's missing value for long-term forecasting. MTCAN model considers the correlation between meteorological factors and PM10 pollutants with large size historical data to improve data quality and long term forecasting performance of PM2.5. MTCAN model includes the TCN model for time series forecasting and the ANN model for fine-tuning prediction results. Fig. 1 displays the graphical illustration of the proposed model framework in detail. Each stage of the proposed model is described in detail in the below subsections.

4.1. Multi-directional imputation

4.1.1. Interpolation at feature matrix level

The data preprocessing of raw air quality datasets is an important step as the quality of the data has a major impact on the forecasting output. Previous research studies evidence indicates a high influence of meteorological pattern and PM10 on PM2.5 forecasting (Liu et al. 2019). Therefore, PM10, meteorological parameters, and their interactions have considered a feature matrix for PM2.5 target variable data imputation and forecasting. Missing values in more than one variable is a serious challenge for environmental data modeling. It is common to have missing values in environmental data, in which case it is not a wise step to consider those partially observed values for analysis. The interpolation technique is adopted to handle the missing values of PM10 and meteorological parameters. It usually considers temporally correlated observations to reconstruct the lost data in the dataset. It has been widely used for real-time air quality dataset for air quality modeling.

Suppose there are a set of features, $(f_1, f_2. \ldots\ldots f_n)$. Among all these features, few of them or all are having missing values. Each features' measurements $f_T = (f_t, f_{t+1}, f_{t+2}. \ldots f_T)$ are having temporal correlation among them so linear interpolation is adopted to recover those missing values of timestamp T. However, interpolation techniques try to reconstruct the missing values by identifying the temporal correlation within a particular data stream but don't capture the temporal correlation among other streams to recover them. So multivariate imputation technique is implemented to impute the target PM2.5 missing values with the interpolated feature matrix.

4.1.2. Multiple imputation for target variable

This research study constructed a multivariate iterative imputer (Bouhlila and Sellaouti 2013; Zhang et al. 2020) to capture temporal correlation among feature matrix and PM2.5 target variables while imputing its missing values. This method preserves the relations in the data and uncertainty about those relations (Buuren and Groothuis-Oudshoorn, 2010). It is a sequential regression multivariate imputation approach which includes a series of univariate models by several iterations instead of just a single large model to impute the target variable. In this case, the PM2.5 pollutant values are imputed by the feature matrix by specifying an imputation model for each feature. PM2.5 value will be regressed by f_1 to f_n values, whose missing values have linearly interpolated initially.

4.2. Temporal convolutional artificial neural network for time series forecasting

Convolutional Neural Network (CNN) has been used from the past decade for analyzing features (Wu et al. 2020a) and visualization tasks (Moor et al. 2019). It is adopted in different domains, like natural language processing and video analytic. However, recent research studies indicate that some new CNN based architectures outperform the traditional Recurrent Neural Network (RNN) structures in machine translation and audio synthesis (Zhang et al. 2019a); Cheema et al. 2018). Encouraged by those applications in various domains, a novel CNN-based architecture is used in this research study for air quality forecasting.

Recently an extended version of CNN, i.e. Temporal Convolutional Network (TCN) (Song et al. 2020; Zhang et al. 2019a) has also been utilized in much research domain and validated its performance. Compared with RNN, CNN is usually used only for feature extraction purposes not for sequential modeling due to lack of memory block to handle long sequences. However, TCN, a variant of CNN outperforms the RNN structures in training efficiency from two aspects. The main components of this model are 1d convolution, causal convolution, dilated convolution, and residual block (Wan et al. 2019). Convolution of TCN can capture local information, so it is better at temporal modeling. The receptive field of dilated causal convolution allows us to capture more input features. It can perform feature learning and predictions in parallel to improve prediction performance. The TCN model has two distinguishing characteristics, which make it different from CNN and RNN structure (Bai et al. 2018).

1. Unlike RNN, it can take input of any length and map it to the same length of the output sequences (Moor et al. 2019). TCN uses Fully Connected Neural Network (FCNN) to achieve this goal.
2. TCN has convolution which is causal in nature so that it can make sure that there is no information leakage from the future to the past. It generates sequential prediction results (Cheema et al. 2018). After each sample is predicted it is fed back into the network to predict the next sample values. As TCN has causal convolution, it does not have recurrent connections so faster to train a longer

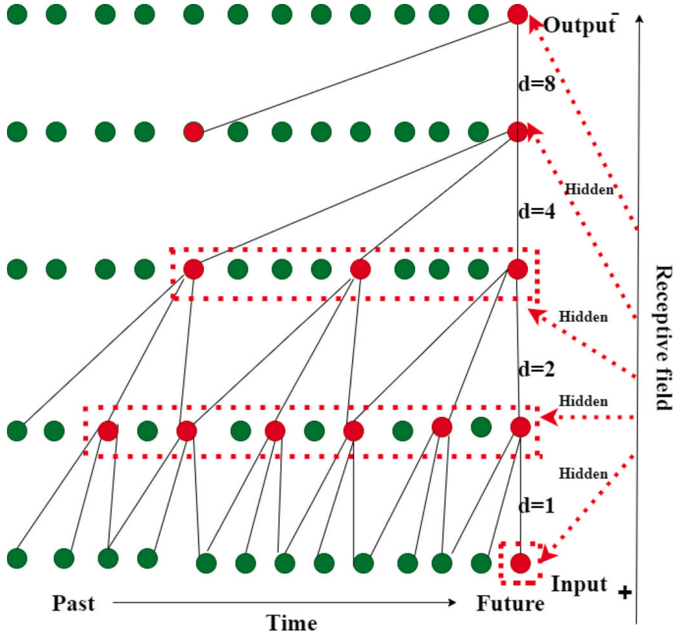

Fig. 2. Dilated causal convolution.

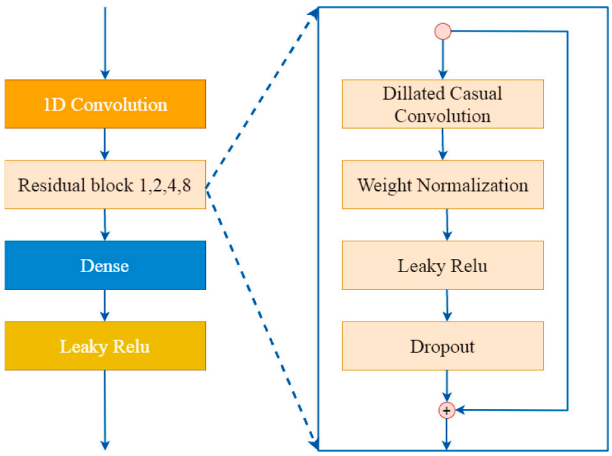

Fig. 3. Residual block.

sequence. In TCN, convolution operations can be performed in parallel, as it uses the same filter size for all the layers. So TCN can be formulated as (Bai et al. 2018),

TCN = 1D FCNN + causal Convolution.

The major drawback of the basic design of causal convolution is that it requires a deep network or large filter size to get a long effective history size (Bai et al. 2018), which does not support basic structures. Larger filter size causes non-convergence issues while a large deeper network causes training issues which ultimately degrade the model prediction performance. To address these limitations, dilated causal convolution is adopted in the TCN structure, where the dilation factor increases with the increasing depth of the network. It has the ability to extend the receptive field. Thus TCN achieves better computational efficiency with the use of a larger receptive field and without the use of larger filter size.

4.2.1. Dilated causal convolution

The convolution operation for input sequence x for the element e can be defined as (Zhang et al. 2019a; Song et al. 2020),

Table 1
Dataset description.

Dataset	CPCB dataset	Urban air quality dataset
Dataset type	Multivariate	Multivariate
Time interval	15 min	60 min
Monitoring site	Talcher, India	Gucheng, Beijing
Sampling period	02/02/2018 to 04/07/2020	01/03/2013 to 28/02/2017
Num of attributes	10	18
Num of site	1	1

$$F(e) = \sum_{m=0}^{k-1} f(m) x_{e-m} \tag{1}$$

The dilated convolution operation for the element e can be defines as,

$$F(e) = \sum_{m=0}^{k-1} f(m) x_{e-d.m} \tag{2}$$

where $f(m)$ is the m^{th} number of filter of its corresponding layer, d is the dilation factor, k represents the filter size and $e - d.\ m$ is the directions of the past. When d becomes 1 then the dilated causal convolution becomes a standard causal convolution. The larger dilated coefficient allows the neurons at the output to represent a wide range of input historical data and also provides flexibility to expand the larger receptive field. The graphical illustration of dilated causal convolution is presented in Fig. 2, with filter size 2 and receptive field size ranging from 1 to 8 (2^0 to 2^3).

4.2.2. *Residual mapping*

With a suitable filter size, it becomes difficult to get accurate prediction accuracy due to training stabilization (Zhang et al. 2019a). Residual mapping operation is followed in the form of a residual block to overcome this type of issue. The residual block is having a shortcut connection to perform the residual mapping from input sequence x to the transformation $f(x)$. The residual block structure of the MTCAN model is presented in Fig. 3.

The residual mapping function can be formulated as (Zhang et al. 2019a; Song et al. 2020),

$$O = \phi(f(x) + x) \tag{3}$$

where ϕ is the nonlinear activation function.

In the proposed architecture, the TCN has two layers of dilated causal convolution and non-linearity within a residual block. Leaky ReLU is used as a nonlinear activation function. The weight normalization process is adopted to convolutional filters for normalization. Besides, the residual block has also a max-pooling layer and dropout layer. Max pooling operations are adopted to reduce the size of the data and computational complexity. The dropout layer with a loss of 0.1 is added after each dilated causal convolution to avoid the overfitting issue while training the model.

4.2.3. *Fine tuning layer*

Fine-tuning is the concept of transfer learning. It adjusts the weights of the previously trained model to improve the prediction accuracy (Lin et al. 2018). We introduce the Artificial Neural Network (ANN) model to fine-tune the forecasting results to optimize the performance. ANN model has a similar concept with the human brain, which has several interconnected neurons acting in parallel. It can be used as a prediction model in the time series domain. ANN model can approximate any function to a desired level of accuracy. Despite network complexity, the ANN model assumes that the predicted and observed values have an underlying relationship between output and input, compared those values, and then updated the weights accordingly to minimize the loss function. In the end, a lower loss function yields better forecasting results.

5. Experiment

5.1. *Experimental datasets description*

In this section, experiments are conducted using two real air quality data sets to evaluate the proposed model performance. Beijing, Multi-Site Air-Quality DataSet collected from the UCI Machine Learning Repository (Chen, 2017) is used as a baseline dataset. Another dataset collected from Central Pollution Control Board, India (CPCB 2018; Madaan et al. 2019; Guttikunda et al. 2019) is also used for model evaluation. A detailed description of the data is shown in Table 1.

5.2. *Experimental setup*

This section describes the software utilized to experiment with the proposed MTCAN and its hyperparameter settings. Keras and

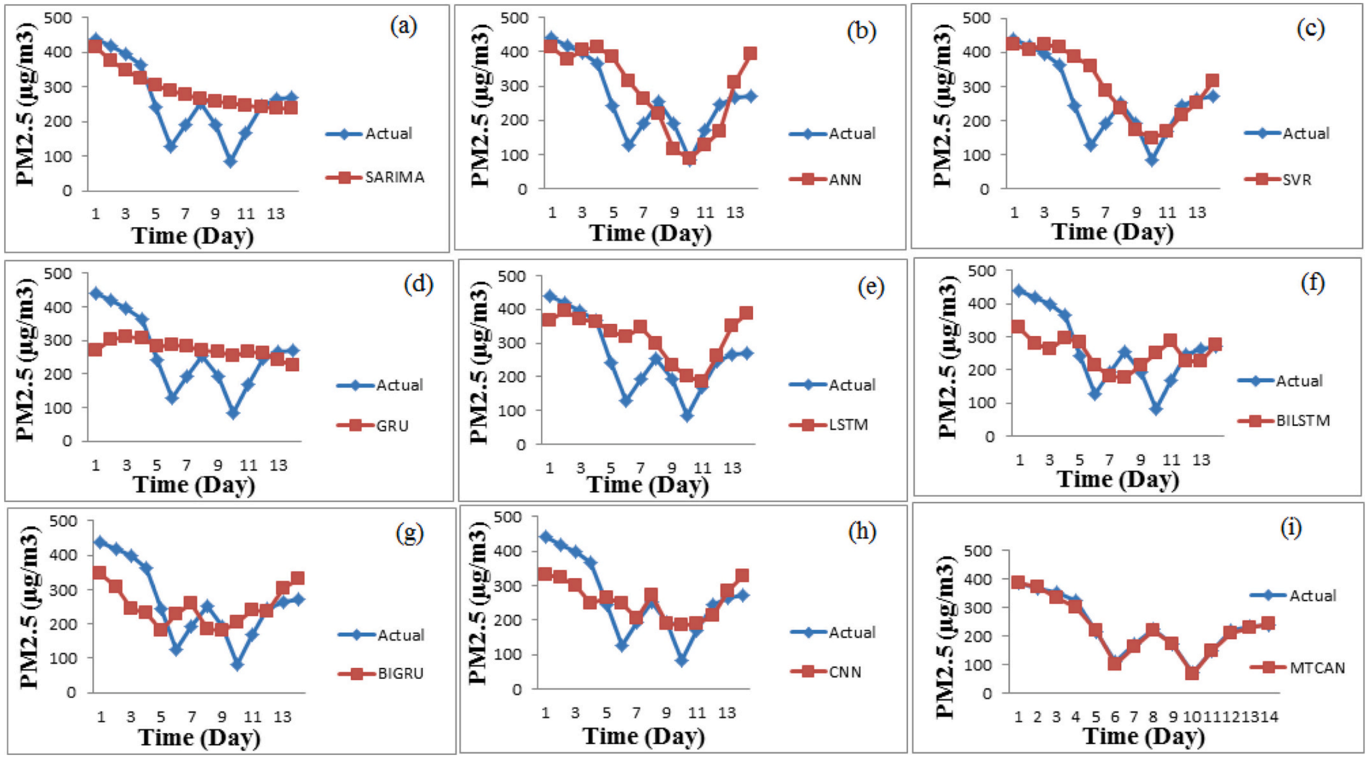

Fig. 4. Forecasting results of MTCAN and state-of-art models for Talcher coal field monitoring site over the next two weeks.

Table 2
Comparison of error metrics of MTCAN and baseline models with CPCB air Quality dataset.

	RMSE							MAE						
Models	3	5	7	9	11	13	14	3	5	7	9	11	13	14
SARIMA	70	65	61	71	77	82	77	70	58	51	57	60	62	60
SVR	92	99	98	99	100	83	82	87	89	83	85	75	56	55
ANN	31	56	99	68	94	86	93	26	54	73	58	81	68	81
CNN	53	71	55	103	88	78	72	44	54	50	83	80	69	58
LSTM	26	71	97	80	129	128	90	24	69	81	52	116	99	71
GRU	88	88	92	77	81	163	97	80	71	71	59	68	144	82
BILSTM	42	155	95	168	135	119	90	33	155	76	143	115	99	74
BIGRU	43	71	52	96	89	92	88	37	64	44	82	75	83	78
MTCAN	11	13	10	15	13	13	9	9	11	8	12	10	10	7

Tensorflow open source libraries are used to develop proposed and other baseline models. The MTCAN model performance is compared with statistical models such as SARIMA, machine learning models such as ANN, SVR, and state of the art deep learning models such as CNN, LSTM, GRU, BILTM, and BIGRU. The baseline models are summarized below,

1. SARIMA, (Samal et al. 2019) a statistical prediction model, is similar to the traditional ARIMA model but usually preferable when data exhibits seasonality.
2. SVR (Murillo-Escobar et al. 2019) is the regression analysis model of Support Vector Machine (SVM) and has a vital role in many time series forecasting applications.
3. ANN (Moghanlo et al., 2021) is a simple neural network model interconnected by several elements. It acquires knowledge by the network through the learning process.
4. CNN (Krizhevsky et al. 2017) has been widely utilized in computer vision and image processing applications. However, one dimensional CNN has been used for sequential modeling tasks such as time series forecasting (Yang et al. 2015). Feature learning and weight sharing are some of the most critical features of CNN.
5. RNN (Cho et al. 2014) is one of the powerful neural network model used for sequential modeling tasks. Though it suffers from vanishing gradient issues, other variants of RNN, such as LSTM (Hochreiter and Schmidhuber 1997) and GRU (Xie et al. 2019) models, are growing its demand as they can solve vanishing gradient issues due to gatting mechanisms.
6. BILSTM (Graves et al. 2005) and BIGRU (Tao et al. 2019) exhibit bidirectional properties of LSTM and GRU models, respectively. These models learn the long-short term dependencies from both the direction to improve the prediction accuracy.

To control the learning process, the hyperparameter setting is an essential task in deep learning modeling. We added dropout with the probability of 0.1 to avoid over-fitting issues. Epochs size is 2000, the batch size is 32, and the prediction zone is set to 14 for long term air quality forecasting. Relu is used as an activation function. RMSProp is chosen as an optimizer for all the models. It considers previous weight updates to update the next weight, instead of considering only current gradient values. So it is usually preferable for sequential modeling tasks and recurrent neural networks. MAE is used as the default loss function for all the models as it can better reflect the actual situation of forecasting error. Moreover, the *Z*-score normalization process is utilized to normalize the data sets and remove the outliers before training all the models. Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) are used as model evaluation indicators presented as below (Du et al. 2018),

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(y_t - \hat{y}_t\right)^2} \tag{4}$$

$$MAE = \frac{1}{n}\sum_{t=1}^{n}\left|y_t - \hat{y}_t\right| \tag{5}$$

where y_t, $\hat{y}_t$ are the ground truth observation and predicted value at t. n presents the total number of samples.

5.3. Long-term forecasting results analysis

The comparative analysis of proposed MTCAN and other state-of-art models are conducted with two real-world datasets. To experiment, each model's forecasting performance, the window size is taken as 30 with different prediction zones (3,5,7,9,11,13,14). The statistical error metrics show that the proposed MTCAN model has superior performance than the other baseline models in RMSE and MAE for all the experimented prediction zones. It is worth noting that though the MTCAN model has excellent performance for each prediction zone and each location, it has the best prediction performance for the next 14 days. It is because the lowers error metrics values signify the better forecasting performance. Though the MTCAN model has superior performance both for the short-term and long-term periods, it works more accurately for the long term period.

Specifically, the PM2.5 forecasting curve for station 1, Talcher coalfield area, is illustrated in Fig. 4. The Fig. 4 shows the forecasting comparison results of eight baseline models and the MTCAN model. The day-wise characteristics of PM2.5 over the 14 days can be

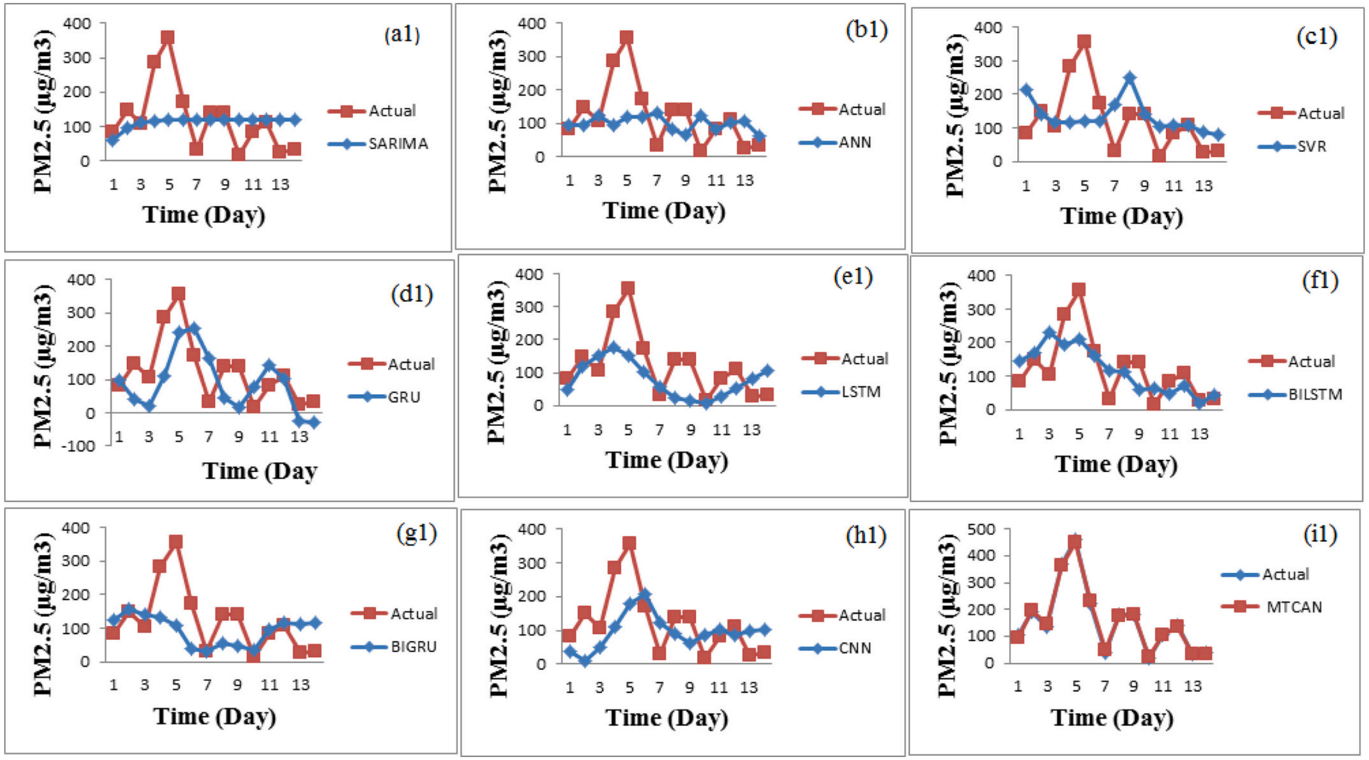

Fig. 5. Forecasting results of MTCAN and state-of-art models for Gucheng monitoring site over the next two weeks.

Table 3
Comparison of error metrics of MTCAN and baseline models with Multi site Air Quality dataset.

	RMSE							MAE						
Models	3	5	7	9	11	13	14	3	5	7	9	11	13	14
SARIMA	74	73	67	75	108	97	94	94	62	62	58	59	72	69
SVR	54	67	52	57	85	107	103	45	57	46	52	66	74	77
ANN	67	52	60	81	114	93	99	56	46	54	68	87	68	73
CNN	22	55	95	98	101	107	92	14	44	89	82	82	78	78
LSTM	84	33	35	58	56	93	87	74	27	29	58	52	72	72
GRU	56	67	78	61	67	99	94	46	55	70	51	57	77	83
BILSTM	62	32	63	72	92	87	70	58	28	47	60	73	64	56
BIGRU	91	72	49	99	93	97	92	81	64	41	78	74	79	78
MTCAN	8	12	10	12	12	13	6	6	9	9	10	9	11	5

observed in Fig. 4. The PM2.5 predicted curve is almost similar to observed values during testing for the proposed model. To further evaluate the model performance, the error metrics statistical table for station 1 is presented in Table 2. Compared to statistical, shallow machine learning and deep learning models, our model has the lowest RMSE of 9 and MAE of 7 for this site, which improves the PM2.5 prediction performance over the 14 days prediction zone. In addition, it is also observed that SARIMA and CNN are the best performing models among all baseline models for long term prediction for this site.

The PM2.5 forecasting results comparison between MTCAN and other models for the second monitoring station, Gucheng, is demonstrated in Fig. 5. The scatters plots of ground truth observations and predicted PM2.5 concentration levels for the entire dataset confirm significantly less difference between the observed and the MTCAN model's predicted values. In contrast, for other models, PM2.5 predicted values are divergent from actual values. Table 3 shows the statistical comparison of error metrics, which indicates that RMSE of 6 and MAE of 5 for the Gucheng monitoring station over the 14 day prediction zone. It can also be seen that the MTCAN model has the lowest error metrics values for both short term and long term prediction zone as the model has the lowest RMSE and MAE values for all 3,5,7,9,11,13 and 14 day prediction zone. Besides, BILSTM and LSTM model are the best working baseline model due to lower RMSE value for Gucheng station.

Moreover, our proposed MTCAN model has the best performance due to its data quality improvement feature and the use of a wide range of historical data distribution for temporal modeling. The fine-tuning layer adds an extra advantage to its prediction results. MTCAN model proves its performance for both short and long prediction zone for both the dataset; it proves its effectiveness in time series forecasting. It is worth noting that the MTCAN model predicted PM2.5 values follow the same trend as observed values for both the monitoring sites. There is very little difference between predicted and observed PM2.5 for all areas. In contrast, other baseline models fail to track the predicted PM2.5 trends for both locations, which indicates their poor forecasting performance. Our proposed model can deal with missing data issues and forecasting tasks simultaneously at the same time, while the other state of art models only deal with the PM2.5 prediction task.

6. Conclusion

This article proposed a novel deep learning architecture that utilizes a novel data imputation approach to impute both meteorological and pollutant values without losing their correlation. The proposed model can use the temporal correlation of meteorological factors and PM10 pollutants to impute PM2.5 missing values. Simultaneously, we considered large size historical data set through the dilated convolution feature of the TCN model for long term temporal forecasting. The results show that the proposed model solves all the discussed issues simultaneously and minimizes prediction error for air quality modeling. The proposed model shows 93% and 90% better accuracy in terms of RMSE values for long-term forecasting than the ordinary GRU model for both the Beijing monitoring site and Talcher monitoring site.

As for future work, we would like to consider traffic and other meteorological factors to perform air quality modeling, which can further improve the forecasting accuracy.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests.

Acknowledgment

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

References

Bai, S., Kolter, J.Z., Koltun, V., 2018. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. arXiv preprint. arXiv:1803.01271.

Bashir, F., Wei, H.L., 2018. Handling missing data in multivariate time series using a vector autoregressive model-imputation (var-im) algorithm. Neurocomputing 276, 23–30.

Batista, G.E., Monard, M.C., et al., 2002. A study of k-nearest neighbour as an imputation method. His 87, 48.

Beretta, L., Santaniello, A., 2016. Nearest neighbor imputation algorithms: a critical evaluation. BMC Med. Informat. Dec. Making 16, 74.

Bouhlila, D.S., Sellaouti, F., 2013. Multiple imputation using chained equations for missing data in timss: a case study. Large-scale Assess. Educat. 1, 4.

Buuren, S.V., Groothuis-Oudshoorn, K., 2010. mice: Multivariate imputation by chained equations in r. J. Stat. Softw. 1–68.

Caillault, É.P., Lefebvre, A., Bigand, A., et al., 2017. Dynamic time warping-based imputation for univariate time series data. Pattern Recogn. Lett. 139, 139–147.

Cheema, N., Hosseini, S., Sprenger, J., Herrmann, E., Du, H., Fischer, K., Slusallek, P., 2018. Dilated temporal fully-convolutional network for semantic segmentation of motion capture data. arXiv preprint. arXiv:1806.09174.

Chen, S.X., 2017. Beijing Multi-Site Air-Quality Data Data Set. https://archive.ics.uci.edu/ml/datasets/Beijing+Multi-Site+Air-Quality+Data.

Chen, S.X., 2018. Beijing multi-site air- quality data data set. URL. https://archive.ics.uci.edu/ml/datasets/Beijing+Multi-Site+Air-Quality+Data.

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y., 2014. Learning phrase representations using rnn encoder-decoder for statistical machine translation. arXiv Preprint. arXiv:1406.1078.

Cobourn, W.G., 2010. An enhanced pm2. 5 air quality forecast model based on nonlinear regression and back-trajectory concentrations. Atmos. Environ. 44, 3015–3023.

CPCB, 2018. Air pollution. URL. http://cpcb.nic.in/air-pollution/.

Dixon, J.K., 1979. Pattern recognition with partly missing data. IEEE Transact. Syst. Man Cybernet. 9, 617–621.

Du, S., Li, T., Yang, Y., Horng, S.J., 2018. Deep air quality forecasting using hybrid deep learning framework. arXiv preprint. arXiv:1812.04783.

Gnauck, A., 2004. Interpolation and approximation of water quality time series and process identification. Anal. Bioanal. Chem. 380, 484–492.

Graves, A., Fernández, S., Schmidhuber, J., 2005. Bidirectional lstm networks for improved phoneme classification and recognition. In: International Conference on Artificial Neural Networks. Springer, pp. 799–804.

Guttikunda, S.K., Nishadh, K., Jawahar, P., 2019. Air pollution knowledge assessments (apna) for 20 indian cities. Urban Clim. 27, 124–141.

Han, Y., Lam, J.C., Li, V.O., Zhang, Q., 2020. A Domain-Specific Bayesian Deep-Learning Approach for Air Pollution Forecast (IEEE Transactions on Big Data).

Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9, 1735–1780.

Junninen, H., Niska, H., Tuppurainen, K., Ruuskanen, J., Kolehmainen, M., 2004. Methods for imputation of missing values in air quality data sets. Atmos. Environ. 38, 2895–2907.

Karmitsa, N., Taheri, S., Bagirov, A., Makinen, P., 2020. Missing value imputation via clusterwise linear regression. IEEE Trans. Knowl. Data Eng.

Kök, İ., Şimşek, M.U., Özdemir, S., 2017. A deep learning model for air quality prediction in smart cities. In: 2017 IEEE international conference on big data (big data). IEEE, pp. 1983–1990.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2017. Imagenet classification with deep convolutional neural networks. Commun. ACM 60, 84–90.

Li, Z., Qin, L., Cheng, H., Zhang, X., Zhou, X., 2015. Trip: an interactive retrieving-inferring data imputation approach. IEEE Trans. Knowl. Data Eng. 27, 2550–2563.

Lin, S.Y., Chiang, C.C., Li, J.B., Hung, Z.S., Chao, K.M., 2018. Dynamic fine-tuning stacked auto-encoder neural network for weather forecast. Futur. Gener. Comput. Syst. 89, 446–454.

Liu, W., Guo, G., Chen, F., Chen, Y., 2019. Meteorological pattern analysis assisted daily pm2. 5 grades prediction using svm optimized by pso algorithm. Atmosph. Pollut. Res. 10, 1482–1491.

Madaan, D., Dua, R., Mukherjee, P., Lall, B., 2019. Vayuanukulani: adaptive memory networks for air pollution forecasting. arXiv Preprint. arXiv:1904.03977.

Malarvizhi, M.R., Thanamani, A.S., 2012. K-nearest neighbor in missing data imputation. Int. J. Eng. Res. Dev. 5, 5–7.

Moghanlo, S., Alavinejad, M., Oskoei, V., Saleh, H.N., Mohammadi, A.A., Mohammadi, H., DerakhshanNejad, Z., 2021. Using artificial neural networks to model the impacts of climate change on dust phenomenon in the zanjan region, north-west iran. Urban Clim. 35, 100750.

Moor, M., Horn, M., Rieck, B., Roqueiro, D., Borgwardt, K., 2019. Temporal convolutional networks and dynamic time warping can drastically improve the early prediction of sepsis. arXiv preprint. arXiv:1902.01659.

Murillo-Escobar, J., Sepulveda-Suescun, J., Correa, M., Orrego-Metaute, D., 2019. Forecasting concentrations of air pollutants using support vector regression improved with particle swarm optimization: case study in aburrá valley, Colombia. Urban Clim. 29, 100473.

Mustafa, Y.T., Tolpekin, V.A., Stein, A., 2011. Application of the expectation maximization algorithm to estimate missing values in gaussian bayesian network modeling for forest growth. IEEE Trans. Geosci. Remote Sens. 50, 1821–1831.

Perez, P., Menares, C., Ramrez, C., 2020. Pm2. 5 forecasting in coyhaique, the most polluted city in the americas. Urban Clim. 32, 100608.

Pozzer, A., Bacer, S., Sappadina, S.D.Z., Predicatori, F., Caleffi, A., 2019. Long-term concentrations of fine particulate matter and impact on human health in Verona, Italy. Atmosph. Pollut. Res. 10, 731–738.

Qin, D., Yu, J., Zou, G., Yong, R., Zhao, Q., Zhang, B., 2019. A novel combined prediction scheme based on cnn and lstm for urban pm 2.5 concentration. IEEE Access 7, 20050–20059.

Quinteros, M.E., Lu, S., Blazquez, C., Cárdenas-R, J.P., Ossa, X., Delgado-Saborit, J.M., Harrison, R.M., Ruiz-Rudolph, P., 2019. Use of data imputation tools to reconstruct incomplete air quality datasets: a case- study in Temuco, Chile. Atmos. Environ. 200, 40–49.

Rantou, K., 2017. Missing Data in Time Series and Imputation Methods. University of the Aegean, Samos.

Rumaling, M.I., Chee, F.P., Dayou, J., Hian Wui Chang, J., Soon Kai Kong, S., Sentian, J., 2020. Missing value imputation for pm 10 concentration in sabah using nearest neighbour method (nnm) and expectation-maximization (em) algorithm. Asian J. Atmos. Environ. (AJAE) 14.

Samal, K.K.R., Babu, K.S., Das, S.K., Acharaya, A., 2019. Time series based air pollution forecasting using sarima and prophet model. In: Proceedings of the 2019 International Conference on Information Technology and Computer Communications, pp. 80–85.

Song, I., Yang, Y., Im, J., Tong, T., Ceylan, H., Cho, I.H., 2019. Impacts of fractional hot-deck imputation on learning and prediction of engineering data. IEEE Trans. Knowl. Data Eng. 32, 2363–2373.

Song, J., Xue, G., Pan, X., Ma, Y., Li, H., 2020. Hourly heat load prediction model based on temporal convolutional neural network. IEEE Access 8, 16726–16741.

Tao, Q., Liu, F., Li, Y., Sidorov, D., 2019. Air pollution forecasting using a deep learning model based on 1d convnets and bidirectional gru. IEEE Access 7, 76690–76698.

Wan, R., Mei, S., Wang, J., Liu, M., Yang, F., 2019. Multivariate temporal convolutional network: a deep neural networks approach for multivariate time series forecasting. Electronics 8, 876.

Wu, P., Sun, J., Chang, X., Zhang, W., Arcucci, R., Guo, Y., Pain, C.C., 2020a. Data-driven reduced order model with temporal convolutional neural network. Comput. Methods Appl. Mech. Eng. 360, 112766.

Wu, R., Zhang, A., Ilyas, I., Rekatsinas, T., 2020b. Attention-based learning for missing data imputation in holoclean. In: Proceedings of Machine Learning and Systems, pp. 307–325.

Xie, H., Ji, L., Wang, Q., Jia, Z., 2019. Research of pm2. 5 prediction system based on cnns-gru in Wuxi urban area. In: IOP Conference Series: Earth and Environmental Science. IOP Publishing, p. 032073.

Xu, X., Yoneda, M., 2019. Multitask Air-Quality Prediction Based on Lstm-Autoencoder Model (IEEE transactions on cybernetics).

Yang, L., Chiang, J.A., 2020. Use case and performance analyses for missing data imputation methods in big data analytics. In: Proceedings of 2020 the 6th International Conference On Computing And Data Engineering, pp. 107–111.

Yang, J., Nguyen, M.N., San, P.P., Li, X., Krishnaswamy, S., 2015. Deep convolutional neural networks on multichannel time series for human activity recognition. In: Ijcai, Buenos Aires, Argentina, pp. 3995–4001.

Yen, N.Y., Chang, J.W., Liao, J.Y., Yong, Y.M., 2019. Analysis of interpolation algorithms for the missing values in iot time series: a case of air quality in Taiwan. J. Supercomput. 1–26.
Yoon, J., Zame, W.R., van der Schaar, M., 2018. Estimating missing data in temporal data streams using multi-directional recurrent neural networks. IEEE Trans. Biomed. Eng. 66, 1477–1490.
Zhang, K., Liu, Z., Zheng, L., 2019a. Short-term prediction of passenger demand in multi-zone level: temporal convolutional neural network with multi-task learning. IEEE Trans. Intell. Transp. Syst. 21, 1480–1490.
Zhang, Y., Wang, Y., Gao, M., Ma, Q., Zhao, J., Zhang, R., Wang, Q., Huang, L., 2019b. A predictive data feature exploration-based air quality prediction approach. IEEE Access 7, 30732–30743.
Zhang, W., Luo, Y., Zhang, Y., Srinivasan, D., 2020. Solargan: multivariate solar data imputation using generative adversarial network. IEEE Transact. Sust. Ener. 12, 743–746.
Zhu, X., Yang, J., Zhang, C., Zhang, S., 2019. Efficient utilization of missing data in cost-sensitive learning. IEEE Trans. Knowl. Data Eng.