EAI Endorsed Transactions
on Scalable Information Systems
Research Article EAI.EU

Auto imputation enabled deep Temporal Convolutional Network (TCN) model for pm2.5 forecasting

K. Krishna Rani Samal[1,*]

[1]Vellore Institute of Technology,Vellore,Tamil Nadu, India

Abstract

Data imputation of missing values is one of the critical issues for data engineering, such as air quality modeling. It is challenging to handle missing pollutant values because they are collected at irregular and different times. Accurate estimation of those missing values is critical for the air pollution prediction task. Effective forecasting is a significant part of air quality modeling for a robust early warning system. This study developed a neural network model, a Temporal Convolutional Network (TCN) with an imputation block (TCN-I), to simultaneously perform data imputation and forecasting tasks. As pollution sensor data suffer from different types of missing values whose causes are varied, TCN attempts to impute those missing values in this study and perform prediction tasks in a single model. The results prove that the TCN-I model outperforms the baseline models.

Keywords: Deep learning; Pollution; Imputation; Forecasting; PM2.5; TCN

Received on 13 February 2024, accepted on 16 June 2024, published on 11 July 2024

doi: 10.4108/eetsis.5102

1. Introduction

High urbanization demands a high quality of life, which ultimately gives rise to traffic emissions and coal emissions. These are becoming significant contributors to increasing air pollution. Around seven million people die worldwide due to hazardous pollutant concentrations every year. Ozone, sulfur dioxide, carbon monoxide, and particulate matter are the major contributors to pollution. However, particulate matter PM2.5 is the critical parameter among them [1]. Particulate matter, PM2.5, became a national crisis worldwide [2] due to its severe harmful impact on both human health and the atmospheric environment. Identifying the air pollution concentration level and forecasting its values are the necessary steps of air pollution management. The governments of polluted countries like China and India have also issued several policies to take the required steps to mitigate its negative impact [3].

Air pollution forecasting has been regarded as an important research topic for researchers. It has a significant r ole i n d isaster p revention a nd ecological decision-making activities. Air pollution forecasting based on historical data demands various methods to provide accurate results [4]. Though time series forecasting is based on past information, a high-quality dataset is required. However, air pollution data concentration varies due to meteorological parameters' impact [5, 6] or may contain missing attributes due to sensor shutdown, natural disasters, or system crashes [7]. So air pollution forecasting method should analyze the climate variables sensitivities and make sure that the source data does not contain missing values. Otherwise, it may mislead into inappropriate decisions. Accurate air pollution forecasting results have a vital role in the early warning system and play a crucial role in environmental and human health protection.

Air pollution modeling is a complex method, as it depends upon several factors like meteorological parameters and traffic emissions. So, the relationship among the parameters should be analyzed before conducting air pollution modeling to identify the required input. Past research studies analyzed from their experimental studies that PM2.5 pollutant is mainly affected by temperature [8, 9], wind direction [10], rainfall [11], wind speed [5, 6], humidity [12, 13], and so on. However, most of the air quality modeling tasks ignore weather conditions for PM2.5 forecasting, which degrades the model performance. This research

* Corresponding author. Email: krishnarani.samal@vit.ac.in

gap motivates researchers to analyze the meteorological factors sensitivity for air quality modeling.

Air pollution modeling is classified into physical-based modeling and data-driven modeling. Physical modeling demands a thorough understanding of environmental science and chemical reactions, whereas data-driven techniques depend upon the dataset's hidden characteristics. Though the physical models [14] are more complex, data-driven approaches [15] are emerging to solve air quality modeling issues in this era of big data.

Recently deep learning-based approaches have increased the demand to handle a vast amount of data. Hence, accuracy can be improved with the excellent quality of the dataset. Nevertheless, due to irregular sampling, natural disaster pollution data may be corrupted, becoming challenging to explore. Sometimes sensor data become critical to analysis due to its manual measurement in most monitoring sites. These unsolved critical issues may bias forecasting results [3, 7, 16].

Hence, the research study developed a neural network model with a deep learning based imputation approach, which can analyze the impact of weather conditions on air pollution and handle the corrupted data simultaneously. The main contributions of this research paper are summarized below,

- In this study, we constructed a TCN-based imputation block that can impute the missing values to better the data quality in the data preprocessing step.
- A new approach is developed to perform PM2.5 forecasting and data imputation in a single run minimizing the training cost. The proposed TCN-I model with dilated causal convolution can effectively perform sequential modeling tasks with huge historical information.
- The proposed model can analyze the climate variables' sensitivity to identify the required input features for better forecasting results.

2. Literature survey

2.1. Air quality forecasting

Samal *et al.* [17] proposed both Seasonal Autoregressive Integrated Moving Average (SARIMA) and FbPROPHET model to forecast SO2, NO2, RSPM, and SPM pollutants using historical pollution information from Central Pollution Control Board. The authors found from their accuracy metric comparison table that though both SARIMA and FbPROPHET have effective air pollution forecasting performance, the FbPROPHET model on log transformation is superior to the SARIMA model. However, the models could not handle the multivariate dataset.

Lagesse *et al.* [18] developed statistical and deep learning prediction models for large office buildings. The authors used various environmental variables as the predictor. The developed models included Multiple Linear Regression (MLR), Distributed Lag Model (DLM), Least Absolute Shrinkage Selector Operator (LASSO), Simple Artificial Neural Networks (ANN), and Long-Short Term Memory (LSTM). The LSTM model outperformed PM2.5 prediction by learning the temporal pattern of predictors. The model is trained and tested for a single building, the model's applicability for other sites and buildings is limited.

S. Freeman *et al.* [19] presents a deep learning model, i.e., Recurrent Neural Network with LSTM model, to predict the hourly concentration of ozone. The authors handled the missing values by first-order differencing of neighboring periods. Among 25 available features, they utilized five features and the LSTM model to improve the prediction accuracy. A decision tree was used to identify the proper input variables. The authors found that removing, minimizing the features, and optimizing the parameters can improve LSTM forecasting model performance. However, the study did not analyze the impact of missing values and their imputation on the overall performance of pollutant forecasting accuracy.

Samal *et al.* [3] identified the correlation between meteorological factors and air pollutants and selected proper input variables to get better PM2.5 forecasting results under weather conditions. The authors experimented with both Odisha, India, and Beijing, China air pollution datasets. The Convolutional LSTM model is utilized for feature extraction and temporal modeling and Sparse Denoising Autoencoder for fine-tuning purposes. The authors compared the proposed model's performance with Support Vector Regression (SVR), ANN [16], Convolutional Neural Network (CNN) [20, 21], LSTM [7, 21], Bidirectional LSTM (BILSTM) [20] and Bidirectional Gated Recurrent Unit (BIGRU) [3] model. The results show that the Convolutional LSTM–SDAE (CLS) model solves most of the multivariate time series problems and has better performance than the other models. However, the model uses the KNN imputation technique, which does not work well for large datasets.

Ge *et al.* [22] developed the Deep Bidirectional and Unidirectional Long Short-Term Memory (DBU-LSTM) model based on the unidirectional and bidirectional properties of RNN. The univariate time series-based prediction models are not efficient enough to analyze the dataset features' hidden correlation, affecting the prediction results. The bidirectional property of the model can easily capture the bidirectional temporal and spatial dependencies from the sequential dataset. The study describes the spatial similarity between different regions. The embedding method is proposed to minimize the dimensionality of the data. The DBU-LSTM model was evaluated with a real-time air quality dataset of Beijing, which outperforms the unidirectional LSTM model. However, the model did not identify the impact of air pollutant influencing factors and their source, which can negatively affect the model's performance.

Yeo *et al.* [23] proposed a Convolutional Neural Network (CNN) model with a Gated Recurrent Unit (GRU) network to perform deep learning based spatiotemporal PM2.5 forecasting. GRU has similar properties to LSTM but without an output gate. It can work on a single point and also on a sequential dataset, so the authors trained the model with meteorological factors and available past PM2.5 data. They could predict the PM2.5 values for the 25 NIER stations in Seoul for 2018. The proposed approach used geographical correlation based on nearby monitoring stations to improve the PM2.5 forecasting accuracy. However, the model performance is limited to PM2.5 pollutants. It became challenging to forecast other pollutants such as PM10 and Ozone due to their different characteristics.

Chen *et al.* [24] proposed an integrated dual LSTM model based on sequence-to-sequence technology, which is used to get prediction values of each component individually. In the second step, a multi-factor forecasting model is developed using LSTM with an attention mechanism. The proposed method used both the weather and spatial features of neighboring sites. Finally, the XGBoosting tree is utilized to integrate these two models, named the integrated dual LSTM model. The model has higher accuracy than Support Vector Regression (SVR), Ridge regression, XGBoost, SLSTM, NLSTM, and single-factor and multi-factor models, but it still has drawbacks. The model gives prediction values with some outliers.

Xie *et al.* [25] analyzed spatial characteristics of PM2.5 monitoring stations and developed CNN's-GRU model to extract spatial features of multi-scale data in a high dimension automatically to develop an advanced PM2.5 prediction framework. The proposed model takes multiple two-dimensional matrices developed with pollutant and climate variables of different monitoring stations in the Wuxi urban area. CNN structures are intended to extract and fuse the data automatically, and then the GRU network is used to capture long-term data variations. The experimental results illustrated that the CNN-GRU model has better forecasting accuracy than Auto-Regressive Integrated Moving Average (ARIMA), LSTM, and GRU networks. This research study fails to include other temporal features to handle temporal dynamics and ignores the seasonal impact on air quality.

Gilik *et al.* [26] constructed a CNN+LSTM-based method to forecast hourly pollutant concentration for several locations efficiently. The authors proposed three methods for performing air quality modeling according to input and output relationships using spatial and temporal features. The first method takes univariate input and provides univariate output; the second method considers multivariate input and generates univariate output. The third method uses multivariate input and gives multivariate output. The proposed methods forecast air pollutant concentrations for multiple locations. The CNN layer is used to extract the relationship between spatial features, whereas the LSTM layer extracts temporal characteristics from the time-series dataset. The study has some limitations as the dataset used has many missing values, so its size for training the model is limited.

Samal *et al.* [27] implemented both deep learning techniques and a geostatistical technique to perform regional forecasting of the study area. The authors developed a CNN-BIGRU-ANN deep learning model and used its prediction results for the available monitoring stations. They then considered these prediction results as input for the ordinary kriging interpolation approach to perform long-term regional forecasting. Though the proposed model has comparatively better results than baseline models, it did not consider the impact of climate variables and data imputation issues of the PM2.5 time series dataset.

2.2. Air quality data imputation

Quinteros *et al.* [28] proposed the use of the multiple imputation method as a tool to reconstruct the missing values of air quality for a mid-size city Temuco, in Chile. The authors stated that multiple imputation techniques could effectively impute the missing attributes of the dataset, for instance, the k-nearest neighbor (KNN) imputation technique. KNN imputation technique reconstructs the missing data accurately while considering spatial covariates but not the entire dataset. The authors found that the lower performance of multiple imputation techniques may be due to incorporated imputation errors. The study has not considered the background values of air pollution, which is the limitation of the study.

Pena *et al.* [29] investigated the pattern of air quality which affects the pollution concentration in Cuenca. The authors tried to repair the continuous missing values of the air quality dataset. They carried out the imputation of missing values of the pollutant time series dataset using Lasso and Ridge regression. These models are evaluated to identify the number of input and output parameters that can be utilized to impute the missing values. The models were based on multivariate linear regression with regularization. Though the model could effectively impute the extreme missing values, the model could not recover the large gap missing values efficiently.

Yen *et al.* [30] aimed to evaluate the interpolation performance of linear regression, SVR, ANN and LSTM model. They obtained the air quality dataset for Taiwan City. The authors presented model-based interpolation techniques to impute the pollutant's missing values. The authors stated that the linear regression and SVR model perform better interpolation during high pollution season with minimal cost. On the other hand, deep learning-based ANN and LSTM networks perform well for the whole year. So the researcher needs to choose suitable interpolation techniques based on their preference and requirements.

Wijesekara *et al.* [31] discussed the six univariate imputation techniques to analyze how they deal with the missing values. They discussed mean, spline, simple moving average, exponentially weighted moving average, and Kalman smoothing models. The comparative analysis shows that the mean imputation works well for a small range of missing values. The imputation performance decreases for spline interpolation with the increase of missing values. The authors proved that the Kalman smoothing technique is the best solution among all the compared models when missing values are random (MCAR) type.

Libasin *et al.* [32] reviewed several single imputation techniques and multiple imputation techniques. They mentioned that single imputation techniques usually ignore the ambiguity and ignore the variance of the data. In contrast, machine learning-based multiple imputation techniques impute the missing values with the best possible estimates. So they suggested applying machine learning based imputation techniques to handle the air pollution attributes.

However, the discussed literature studies either focus on data imputation or forecasting but can not solve both issues simultaneously in a single process. Therefore, this research study constructed a model that can perform both the data imputation and forecasting tasks in a single model, which provides more accurate prediction results. In addition, the study compared the performance of the proposed model with the baseline SARIMA, FbPROPHET, ANN, SVR, CNN, LSTM, GRU, BILSTM, and BIGRU model to identify the effectiveness of the proposed model as compared to others.

3. Problem statement

Several time series data generated from the X sensor stations can be represented as, $V = \{v_1, v_2, v_3, ...v_T\}^T \in R^{T\times X}$ with T time steps. Each vector $v_t \in R^X$ represents the X sensors' air pollution level at time t. Each sensor X gives both meteorological parameters information with air pollution concentration level. In reality, each air pollution sensor generates missing attributes due to sensor failure or may be due to some natural disasters. So the collected dataset may have corrupted data. To deal with these missing values, a masking vector is required to identify whether the data has corrupted values or not. Therefore, an imputation block is required to attribute those missing sensor values.

In this research study, air pollution prediction model with an imputation block aims to learn a function $f(.)$ to map a T step of historical air pollution dataset to the next subsequent step of air pollution state, which can be described as, $f([v_1, v_2, v_3....v_T]) = v_{T+1}$. This a multivariate regression problem, that tries to minimize the loss function for accurate forecasting results.

4. Data collection

In the past decades, China has undergone urbanization, which demands energy consumption and industrialization. Residential energy is one of the major contributors to PM2.5 in China [33], which has the highest impact on premature death in China. Land traffic is another major contributor to PM2.5 pollution in China. Mortality attributable to PM2.5 in China is much higher than road accidents and other death causes. Chronic health and premature death mortality are uncertain due to increased pollution levels because of increased industrialization, and traffic emissions.

Due to rapid economic growth in China [34], the air pollution level in Beijing city became prominent. Beijing is one the most developed cities in China and has been accompanied by intense air pollution concentration. Keeping these aspects of air pollution into consideration, Beijing is considered the study area.

The air pollution dataset for Beijing [35] is collected for experimental purposes. The dataset [36] contains pollutant and meteorological parameters features from the year 2013 to 2017. The daily 24-h average PM2.5 pollutant concentration is used to forecast its values in the next 14 days.

The detailed description of the dataset is presented in Table 1.

Table 1. Dataset description

UCI Machine learning repository dataset	Description
Characteristics	Multivariate
Time span	60 minutes
Location	Gucheng, Beijing
Sampling period	01/03/2013 to 28/02/2017
Number of parameters	18
Number of monitoring locations	12
Total Number of instances	420768

5. Proposed methodology framework

The proposed architecture is presented in Figure 1. Once the data is collected, the data preprocessing step is taken to get better quality data. In the first step of the data preprocessing, the model identifies the essential input features and then imputes the missing values of the dataset using TCN based imputation block. After data preprocessing, the TCN model performs time series forecasting of PM2.5 pollutant. Each part of the data preprocessing step and proposed TCN with imputation block (TCN-I) architecture is described below.

5.1. Data preprocessing step

Input parameter selection. It is essential to identify the required input features in the data preprocessing step before the air quality modeling step. The objective of feature selection is to remove the irrelevant input features and keep only the necessary inputs. This research study calculated the Pearson Correlation Coefficient (R) to identify the temporal correlation between climate variables and PM2.5 pollutants to find out the vital input features for PM2.5 modeling as shown in Table 2.

Suppose a and b are the time series vector for PM2.5 and climate variables, then R can be computed as,

$$R = \frac{n\sum_{m=1}^{n} a_m b_m - \sum_{m=1}^{n} a_m \sum_{m=1}^{n} b_m}{\sqrt{n\sum_{m=1}^{n} {a_m}^2 - (\sum_{m=1}^{n} a_m)^2}\sqrt{n\sum_{m=1}^{n} {b_m}^2 - (\sum_{m=1}^{n} b_m)^2}} \tag{1}$$

If the R value is between 0 and 1, then the variables a and b have a positive correlation. If the value lies between -1 to 0, then there exists a negative correlation between the two variables.

Table 2 shows that the PM2.5 pollutant has a positive correlation with wind direction and dew point and a negative correlation with other parameters. It can be observed from Table 2 that all the parameters are poorly correlated with PM2.5 pollutants, so all the variables can be taken as input features directly.

Handling missing values. PM2.5 forecasting models based on data-driven techniques have excellent performance in environmental data engineering. A large number of models have been proposed to perform PM2.5 forecasting. Air quality modeling studies focus on improving prediction accuracy. Air quality modeling based on sensor data is complex due to its characteristics. Firstly sensor data are multivariate datasets, and second, they face missing data values. Various methods have been developed to handle those missing values. However, conducting data imputation and data forecasting tasks individually may lead to unnecessary prediction errors. Solving the data imputation and data forecasting is a two-step process where it is difficult to recognize the missing data patterns, which generates a bias in forecasting results.

In real-time air pollution forecasting tasks, it is challenging to perform data imputations. It is costly due to both individual training processes. Hence looking into this issue, TCN is used to perform both data imputation and data prediction tasks, which can easily handle the long-term temporal dependencies that exist in the dataset without causing data exploding and data vanishing issues. As the TCN model can effectively handle huge past information and identify the long-term temporal dependencies, TCN-based data imputation block can easily handle the unevenly spaced missing value in sequential data to achieve the objective.

One easy option to deal with the missing values before the data prediction task is to ignore those missing values. The other method is to work on the data imputation approach to deal with those values. Several univariate and multivariate imputation methods exist, which can easily impute the missing value but fail to perform the data prediction task. Among all the data imputation tasks, linear, mean, median interpolation, and multivariate imputation by chained equations (MICE) [37] methods are statistical imputation techniques. Miss forest [38], K nearest neighbor (KNN) [39, 40], Expectation-Maximization (EM) algorithm [41], and matrix factorization are machine learning-based imputation approach.

As time-series dataset has long-term dependencies, several deep learning imputation techniques are also proposed to capture this while conducting data imputation, such as Deep learning model based on GRU (GRU-D) [42], Multiple Imputation Using Denoising Autoencoders (MIDA) [43], Multi-directional Temporal Convolutional Artificial Neural Network (MTCAN) [16], Temporal Convolutional Denoising Autoencoder (TCDA) [7], etc. However, the recurrent-based methods require a considerable amount of memory and training time as they follow backpropagation through time to train the model. So this research study developed a TCN-based data imputation block, which can provide data imputation and handle a substantial multivariate dataset and its long-term dependencies to perform forecasting with less memory and training time requirement.

Once, the missing values of PM2.5 pollutants are imputed, the Z score normalization approach is adopted to remove the outliers from the dataset. Sometimes few variables have relatively greater values and a wider range of values, negatively affecting prediction results. To solve this issue scope of

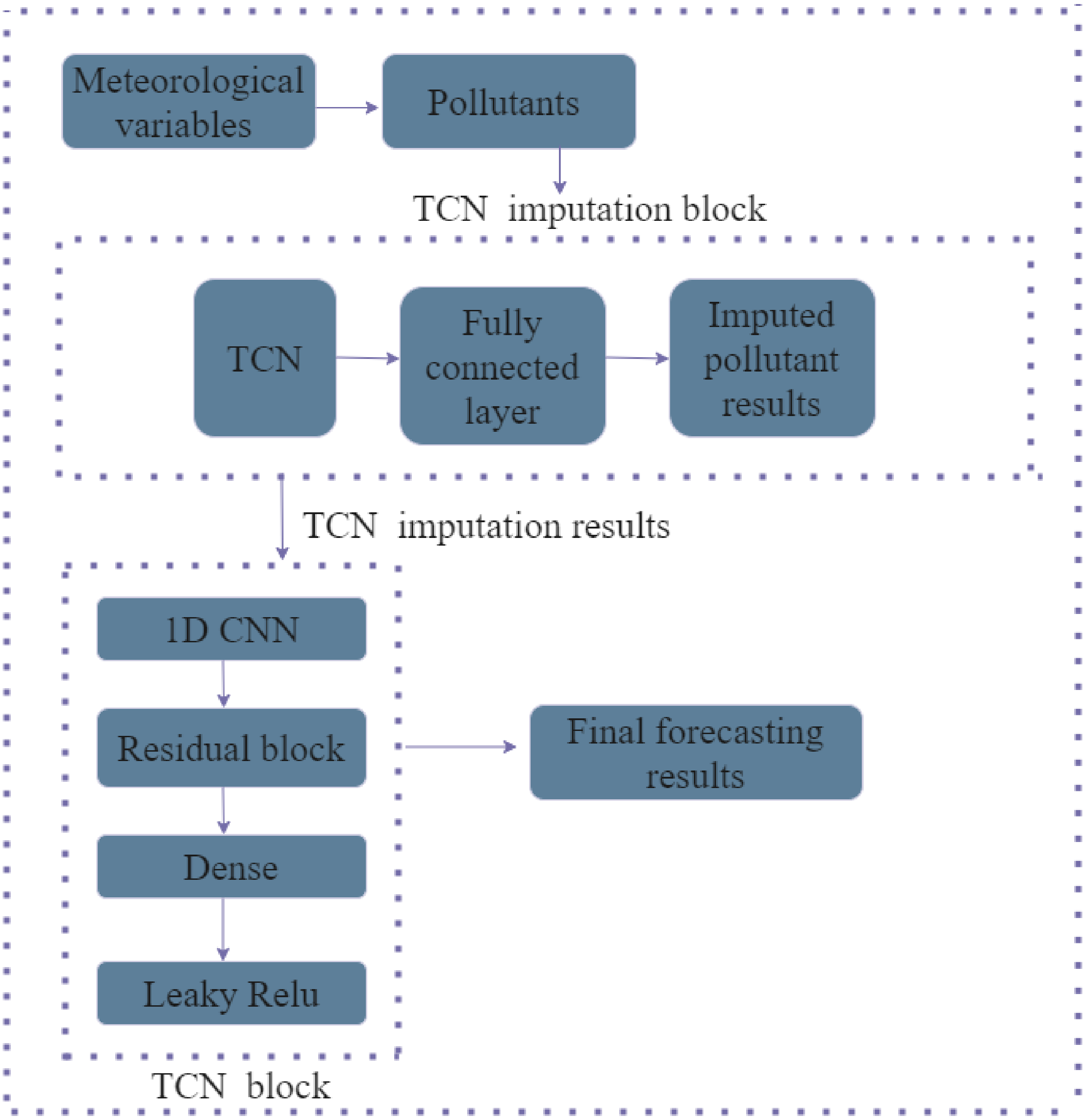

Figure 1. Proposed TCN-I architecture.

the variables should be adjusted through Zscore normalization.

Zscore normalization can be computed as below, [44],

$$X = \frac{y - \mu}{\sigma} \tag{2}$$

where X is represents the Z-score standardized value and y represents the concentration value of the element, μ and σ denote the mean and standard deviation, respectively.

5.2. Time series forecasting

After input feature selection and TCN based data imputation, the TCN network performs temporal modeling to generate the PM2.5 prediction results. The detailed structure of the TCN block is described in the below subsections.

Temporal convolutional network. Compared to LSTM and GRU, the TCN model [45] has a better memory for sequential modeling tasks and performs large-scale parallel processing. It has improved in the following ways,

Table 2. Correlation Coefficient values.

R	PM2.5	Wind speed	Wind direction	Atmospheric temperature	Pressure	Dew point
PM2.5	1	-0.27	0.38	-0.08	-0.02	0.14
Wind speed	-0.27	1	-0.43	-0.21	0.21	41
Wind direction	0.38	-0.43	1	0.06	-0.07	0.20
Atmospheric temperature	-0.08	-0.21	0.06	1	-0.86	0.90
Pressure	-0.02	0.21	-0.07	-0.86	1	-0.80
Dew point	0.14	-0.33	0.20	0.90	-0.80	1

- TCN has causal convolution, which is utilized to better handling the sequential dataset.

- TCN utilizes dilated causal convolution and residual mapping concepts to remember a huge amount of historical information [46].

Causal convolution. Causal convolution mainly assumes that the predicted output depends upon only the current layer elements and past layers' results, not the future inputs. The standard convolution output depends upon the future input at any instant of time; it may or may not rely upon past information, which is irrelevant for sequential modeling tasks. However, the causal convolution relies entirely on the historical time series dataset.

To make sure that the output has the same length as the input, TCN utilizes one dimensional fully connected network [47].

Dilated convolution. The reverse training issue of deep learning networks can be solved by using Dilated convolution. A simple causal convolution looks at the historical information linearly, challenging to analyze huge historical details to apply for a sequential task. To solve the issue, TCN used dilated convolution [48].

Dilated convolution is a causal convolution, where a filter is applied over an area larger than its length by skipping some input values with a certain step to effectively allow the network to have a vast receptive field with just a few layers. The Dilated Causal Convolution with dilation factor 4 is illustrated in Figure 2.

The convolution operation for sequence a for the element b can be presented as [16],

$$F\,(b) = \sum_{n=0}^{k-1} f\,(n) a_{b-n} \tag{3}$$

Similarly, the dilated convolution operation for the element b can be represented as,

$$F\,(b) = \sum_{n=0}^{k-1} f\,(n) a_{b-d.n} \tag{4}$$

where $f(n)$ presented the n^{th} number of filter of its corresponding layer, d denotes the dilation factor, k denotes the filter size and $b - d.n$ denotes the directions of the past. when d value becomes 1 then the dilated causal convolution is converted into a standard causal convolution.

Receptive mapping. It is very critical to achieve good prediction accuracy with a suitable filter size. So TCN adopted residual mapping, a jumper connection for residual convolution. The residual block has a shortcut jumper connection to conduct a residual mapping from input sequence y to the transformation $f(y)$. The residual mapping operation can be defined as [7],

$$O = \phi(F(y) + y) \tag{5}$$

Where ϕ represents a nonlinear activation function.

The utilized TCN block does not have a recurrent connection; thus, it does not use the backpropagation through time to train. They can be trained in parallel for a faster training process and optimize GPU usage. The TCN model does not exhibit the exploding gradients and vanishing gradients issues and can learn from past data without any problems.

6. Experimental setting

Hyperparameter setting is an essential step in training the deep learning models. All the deep learning models are developed using the Google Colab environment. This research study adopted 80 % of data as training, 10 % as the validation set, and the remaining 10 % as a test dataset to train the model. The trial and error method [49–51] is utilized to identify the

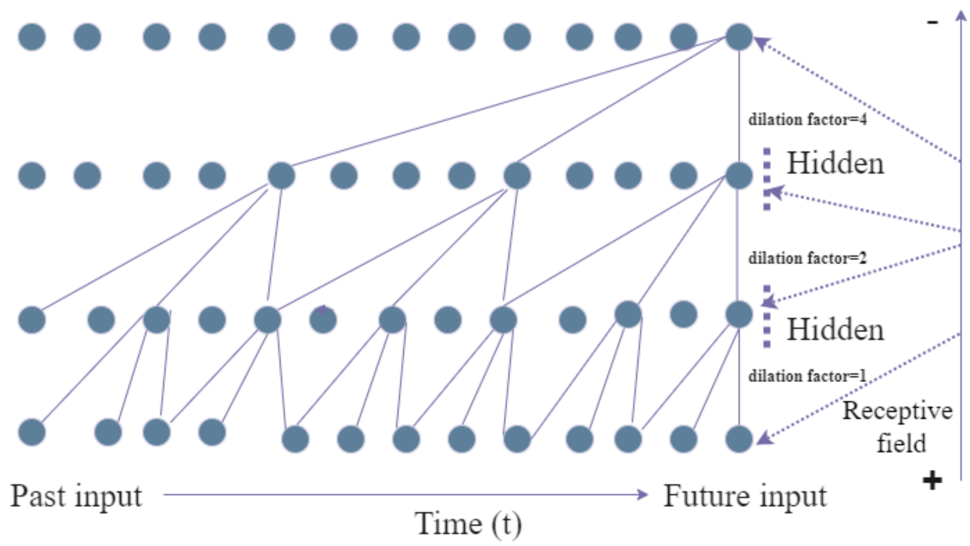

Figure 2. Dilated Causal Convolution with dilation factor 4 and filter size (k)=3.

hyperparameter setting. The proposed model uses two layers of TCN, where the first layer of TCN has 64 filters and the second TCN layer has 128 filters, with a dilated factor of size 4. The proposed model uses three dense layers, each of them having 50 nodes. This research study utilized RMSProp as an optimizer and Mean Square Error (MAE) as a loss function.

The learning rate is a hyperparameter for parameter optimization of the model. It identifies the step size at each iteration towards minimizing the loss function. Usually, its values lie between the interval (0,1). The amount of the weights is updated while training is known as step size or learning rate. We have used the learning rate value as 0.02. Activation functions are the mathematical functions used to determine the output of the neuron. Relu is adopted as an activation function for the experiment. Epoch is set as 2000. The batch size is set as 16. The experimental settings of baseline models are presented in Table 3. Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE) are used as evaluation metrics to compare the performance of all the baseline models and the proposed model. The error metrics can be computed as,

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(z_t - \hat{z}_t)^2} \tag{6}$$

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|z_t - \hat{z}_t| \tag{7}$$

$$MAPE = \frac{\sum_{t=1}^{n}\left|\frac{z_t - \hat{z}_t}{z_t}\right|}{n} \times 10 \tag{8}$$

where z_t, $\hat{z}_t$ are the original observation and forecasting value at t. n presents the total number of samples.

RMSE represents the square root of Mean Square Error (MSE) values. It is used to measure the standard deviation of residuals. MSE is the average of the squared difference b etween t he o bservation a nd predicted values in the dataset. It is used to compute the variance of the residuals. As the name suggests, MAE is the mean of the absolute error, i.e. mean of the difference between predicted and actual values. MAPE represents the average of the percentage errors.

7. Comparison of forecasting models

Before training the proposed model, the data preprocessing step is conducted. The correlation coefficient is computed in the data preprocessing step to identify the required input features to train the model, as shown in Table 2. Different combinations of meteorological parameters with PM2.5 pollutants are trained with the proposed TCN-I model, and the forecasting results are illustrated in Figure 3-10. It can be easily understood from the forecasting results that the PM2.5 predicted values are more accurate when trained with all the meteorological features, as shown in Figure 10.

To further compare the forecasting performance of the proposed TCN-I model with different climate variables, RMSE, MAE, and MAPE values are computed as shown in Table 4. The comparative analysis, demonstrated in Table 4 proves that the proposed model had better accuracy when trained with the combination of all features, i.e., with all the input meteorological parameters and PM2.5 pollutant due to the lowest error metrics values. With the variety of all

Table 3. Hyperparameter setting

CNN	Convolution layer=2, Filter size=128, Kernel size=2, Activation function=Relu, Maxpooling operation with pool size 2, Dropout=0.2, Epochs=2000, Optimizer = RMSProp,Loss function=MAE
LSTM	LSTM layer=2, First LSTM layer has 100 LSTM unit, Second layer has 50 LSTM unit, Activation function=Relu, Maxpooling operation with pool size 2, Dropout =0.2, Epochs=2000, Optimizer=RMSProp, Loss function=MAE
GRU	GRU layer=2, First GRU layer has 100 GRU unit, Second layer has 50 GRU unit, Activation function=Relu,Maxpooling operation with pool size 2, Dropout =0.2, Epochs=2000, Optimizer=RMSProp, Loss function=MAE
BILSTM	BILSTM layer=2, First BILSTM layer has 100 LSTM unit, Second layer has 50 LSTM unit, Activation function=Relu, Maxpooling operation with pool size 2, Dropout =0.2, Epochs=2000, Optimizer=RMSProp, Loss function= MAE
BIGRU	BIGRU layer=2, First BIGRU layer has 100 GRU unit, Second layer has 50 GRU unit, Activation function=Relu,Maxpooling operation with pool size 2, Dropout =0.2, Epochs=2000, Optimizer=RMSProp, Loss function=MAE

input features, RMSE was reduced to 7, MAE to 6, and MAPE to 11. So all the meteorological parameters are considered to perform PM2.5 forecasting.

In this section, the forecasting performance of the proposed TCN-I model is compared with statistical models like SARIMA and FbPROPHET, machine

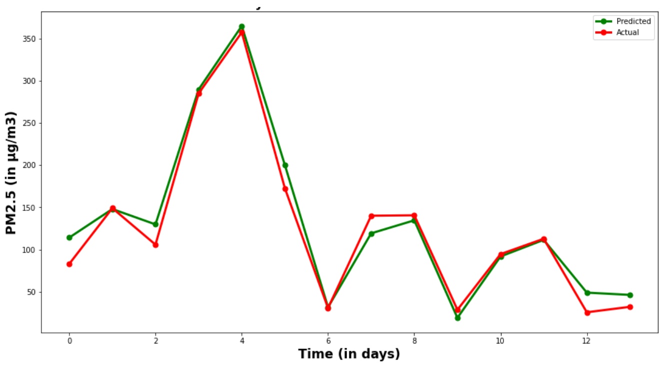

Figure 3. PM2.5 prediction result with temperature parameter.

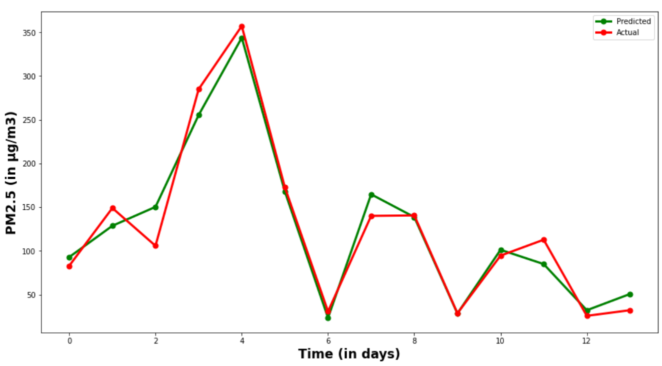

Figure 4. PM2.5 prediction result with pressure parameter.

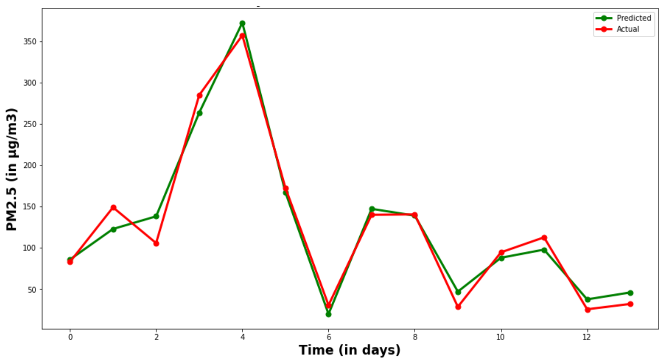

Figure 5. PM2.5 prediction result with dew point parameter.

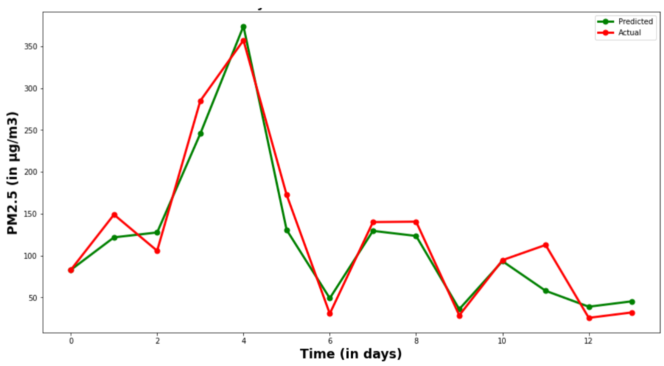

Figure 6. PM2.5 prediction result with wind speed parameter.

learning models like SVR and ANN, deep learning models like CNN, LSTM, GRU, BILSTM, and BIGRU networks. The experiment is conducted with the benchmark dataset collected from the UCI machine learning repository. The baseline models are described as below,

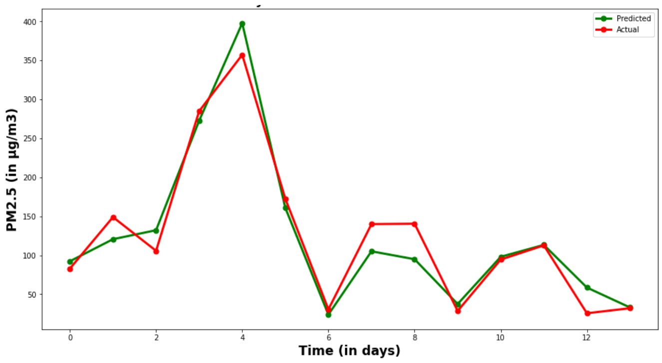

Figure 7. PM2.5 prediction result with wind direction parameter.

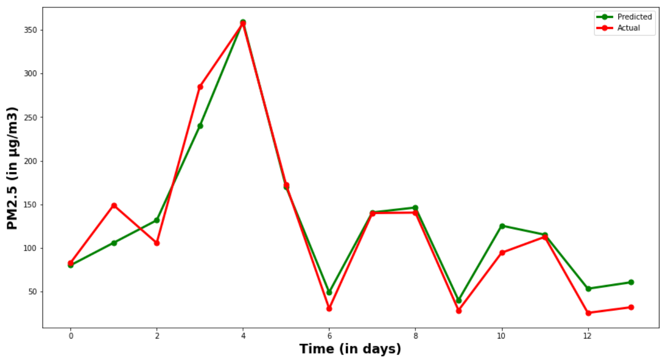

Figure 8. PM2.5 prediction result with rainfall parameter.

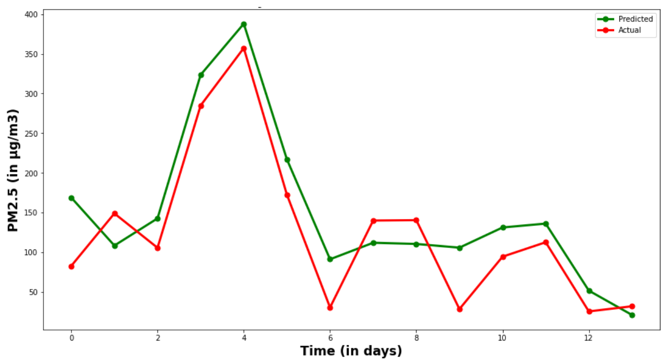

Figure 9. PM2.5 prediction result without meteorological parameter.

- Seasonal Autoregressive Integrated Moving Average (SARIMA) [16] is a statistical forecasting model and based on seasonal trends.

- FbPROPHET [16, 17, 37] is a recently developed prediction model, which is mainly based on trend, seasonality, and holidays.

- SVR is a regression part of the Support Vector Machine model primarily used to perform regression tasks [52].

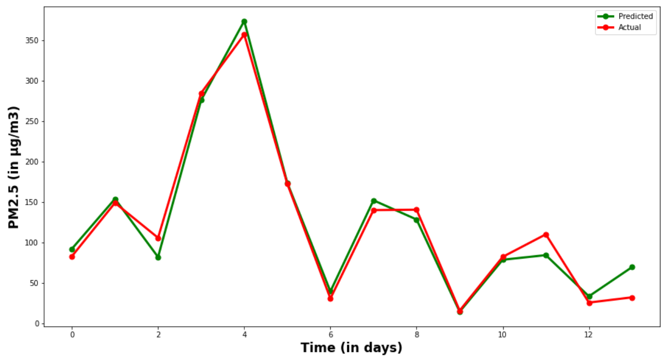

Figure 10. PM2.5 prediction result with all features.

- ANN [16, 53] model interconnects the nodes and finds the interaction among them through the neuron and works like a human brain to generate outputs.

- One dimensional CNN [16] is used to analyze a one-dimensional time series dataset. It can reduce the complexity of huge datasets by reducing their dimensionality. Weight sharing and data sparsity are two important features of CNN to handle a vast dataset [52].

- LSTM network and GRU are two significant categories of Recurrent Neural Network (RNN) networks. They are usually used to handle the long-term dependencies of past historical information to perform temporal modeling with the help of their gating mechanism [52].

- Bidirectional LSTM (BILSTM) [21, 54] and Bidirectional GRU (BIGRU) [16, 21, 55] are two variants of LSTM and GRU structure respectively, which can analyze the dataset in both forward and backward direction to identify the long-short term dependencies of the dataset very effectively to perform time series forecasting.

Table 5 shows the comparison of prediction error metrics of the proposed TCN-I model and baseline models for the prediction horizon of 14 days. Table 5

Table 4. PM2.5 forecasting with different climate variables.

Variables	RMSE	MAE	MAPE
Temperature+PM2.5	16	12	19
Pressure+PM2.5	19	15	17
Dewpoint+PM2.5	15	13	20
WS+PM2.5	25	20	23
Rainfall+PM2.5	23	17	19
PM2.5	45	40	65
WD+PM2.5	23	18	23
All features	7	6	11

indicates that the LSTM model has the worst forecasting performance among all the models, and RMSE is 137. Statistical SARIMA and FbPROPHET model has better performance than machine learning based ANN and SVR models in terms of RMSE. Though the statistical models, i.e., SARIMA and FbPROPHET, have better prediction performance, these univariate models can not handle the multivariate properties and uncertainty of a huge dataset. RMSE values of deep learning-based CNN, GRU, BILSTM, BIGRU and the proposed TCN-I models are 95, 97, 98, 91, and 7. It was evident that the proposed TCN-I model has the lowest error metrics values (RMSE, MAE, and MAPE) as the predicted values are nearly equal to the original values for the proposed model. The TCN-I model has 94% better accurate forecasting performance than the LSTM model in terms of RMSE.

To further analyze and compare the forecasting effect of the proposed TCN-I and baseline models, scatter plot results are presented in Figure 11-20. The X-axis in the figures represents the timestamp in days, and the Y-axis represents the PM2.5 concentration values. The red color line shows the actual PM2.5 values, whereas the green color line represents the predicted PM2.5 values. Figure 20 shows that the PM2.5 predicted values of TCN-I and actual values are almost the same and have less difference between them than other baseline models.

Figure 11-20 shows that the BILSTM model has better PM2.5 prediction results than the LSTM model. BILSTM can process the data in forward and backward directions and analyze the hidden temporal pattern more effectively than the LSTM model. Therefore, the BILSTM model forecasting curve is much closer to the original values than the LSTM model. In addition to that, it can be seen from the figures that the GRU model has better forecasting results than the LSTM model, as the GRU model's 14 steps ahead predicted values and observed values have less difference than the LSTM model. Figure 11-20 shows that all models' predicted values do not follow the fluctuations of actual PM2.5 values, except the developed model TCN-I. Therefore TCN-I has the best forecasting performance over the 14 days.

8. Comparison of imputation performance

In this section, the proposed model's imputation performance is compared with other baseline imputation approaches. The TCN-I model's imputation performance is compared with the cubic, linear, spline interpolation, mean, median, KNN, MICE, and LSTM imputation techniques with the TCN model for forecasting tasks. The experimental results are presented in Table 6-8. The PM2.5 forecasting performance is compared with random missing values at a missing rate ranging

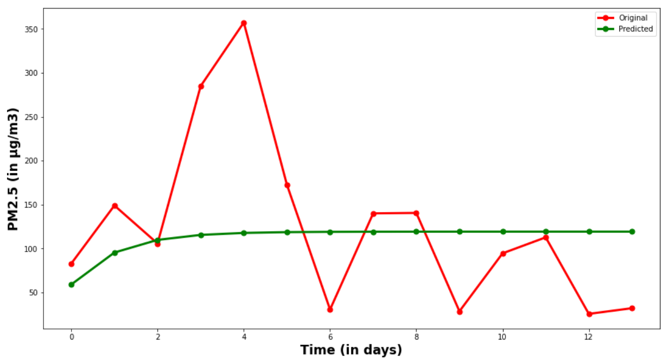

Figure 11. PM2.5 prediction result of SARIMA model.

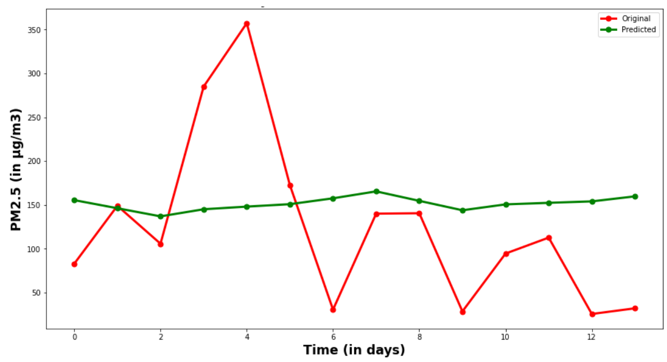

Figure 12. PM2.5 prediction result of FbPROPHET model.

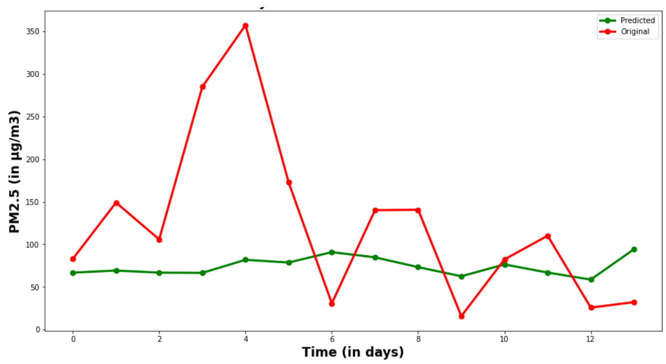

Figure 13. PM2.5 prediction result of ANN model.

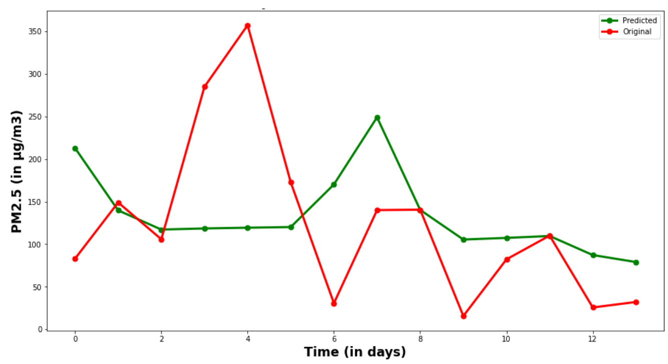

Figure 14. PM2.5 prediction result of SVR model.

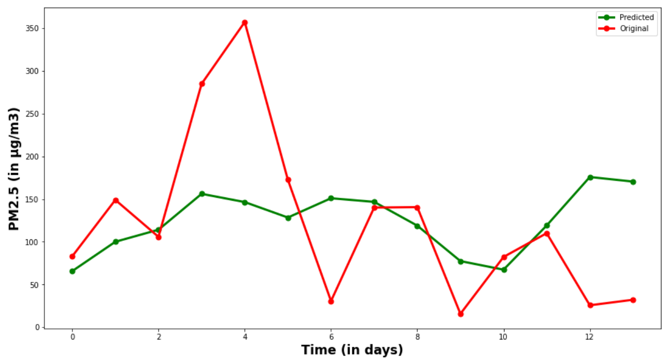

Figure 15. PM2.5 prediction result of CNN model.

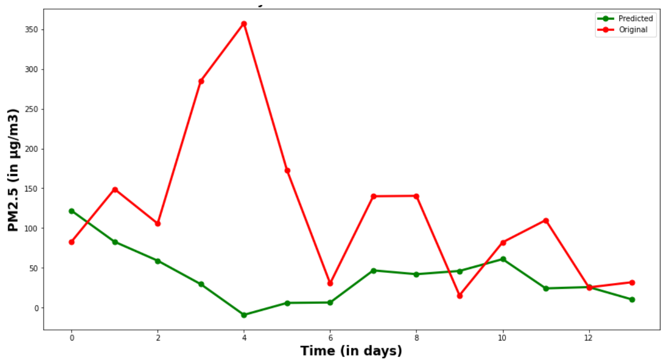

Figure 16. PM2.5 prediction result of LSTM model.

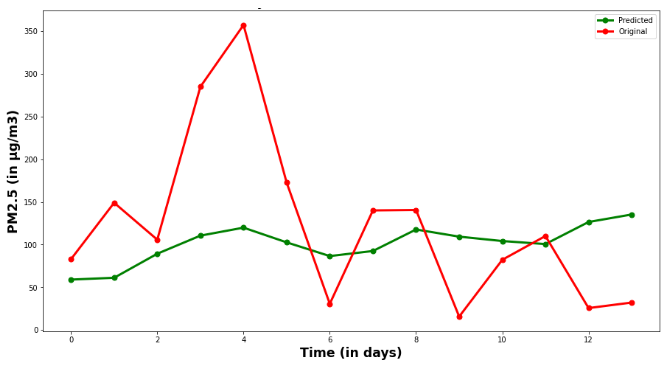

Figure 17. PM2.5 prediction result of BILSTM model.

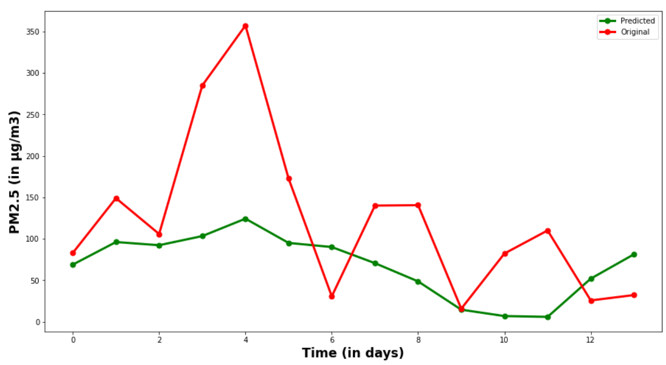

Figure 18. PM2.5 prediction result of GRU model.

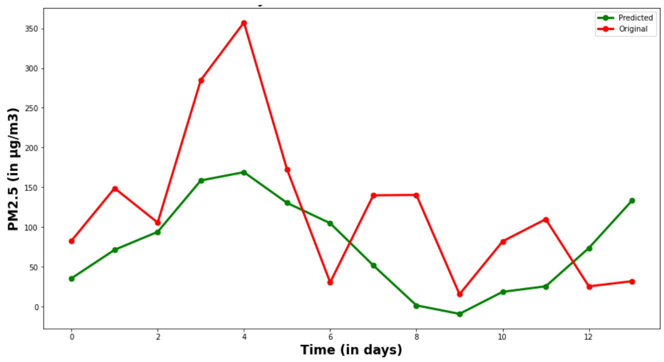

Figure 19. PM2.5 prediction result of BIGRU model.

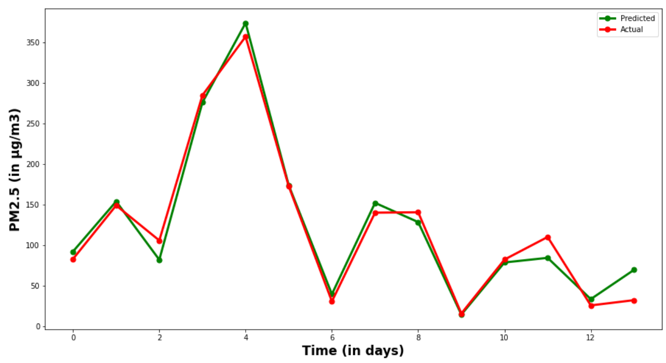

Figure 20. PM2.5 prediction result of TCN-I model.

from 20% to 50%. The proposed model outperforms the other models in both cases when the missing rate is low and high due to its lowest RMSE, MAE, and MAPE values. Lowering the error metrics values better the model performance. The TCN-I model works better at a high missing rate than the other baseline models. Suppose the information of the observed data decreases, the disappeared rate increases. To handle this situation, data imputation quality should be high. The results

Table 5. Prediction error metrics comparison over 14 days.

Models	RMSE	MAE	MAPE
SARIMA	94	69	109
FbPROPHET	98	78	147
ANN	107	78	90
SVR	103	77	132
CNN	95	70	145
LSTM	137	93	71
GRU	97	74	71
BILSTM	98	76	132
BIGRU	91	79	104
TCN-I	7	6	11

Table 6. Prediction error metrics (RMSE) with random missing values.

Model/Missing rate(%)	20	30	40	50
Linear-TCN	91	86	85	71
MICE-TCN	73	77	68	69
Spline-TCN	49	114	121	71
LSTM-TCN	134	98	110	80
Mean-TCN	119	130	119	129
Cubic-TCN	117	79	135	110
Median-TCN	67	121	67	59
KNN-TCN	63	76	62	73
TCN-I	26	28	27	29

Table 7. Prediction error metrics (MAE) with random missing values.

Model/Missing rate(%)	20	30	40	50
Linear-TCN	63	69	68	57
MICE-TCN	59	67	48	52
Spline-TCN	41	81	105	72
LSTM-TCN	98	74	110	66
Mean-TCN	91	95	87	91
Cubic-TCN	91	62	113	91
Median-TCN	52	85	54	43
KNN-TCN	53	54	40	58
TCN-I	21	22	21	23

Table 8. Prediction error metrics (MAPE) with random missing values.

Model/Missing rate(%)	20	30	40	50
Linear-TCN	143	79	85	105
MICE-TCN	125	121	109	110
Spline-TCN	13	23	45	20
LSTM-TCN	115	114	186	144
Mean-TCN	98	98	149	98
Cubic-TCN	40	28	46	36
Median-TCN	71	94	94	58
KNN-TCN	67	67	68	79
TCN-I	38	24	25	45

show that the TCN-I model provides better imputation quality. The results also illustrated that the TCN-I model with the imputation block indirectly helps to improve the forecasting results.

The government of China has carried out some initiative plans, and emphasis has been given to day-to-day activities in urban cities to minimize the pollutants' concentration level [3]. Intensive air quality measurements are required for proper air pollution control, where air pollution prediction results play a crucial role. The prediction results can be utilized for traffic mo nitoring st ations or industrialization locations, which are the primary source of pollutants to take preventive steps against future pollution levels.

9. Conclusion

In this research study, we reformulate the TCN approach to perform the data imputation and prediction task. The proposed TCN-I model has an extra imputation block to handle the missing values. The proposed TCN-I prediction model can identify the appropriate input parameters in the data preprocessing step to achieve the PM2.5 forecasting task. After input feature selection, the model imputes the missing values for a better quality dataset using a TCN-based imputation block. Then the preprocessed data is used as input for the proposed model with dilated causal convolution, having a dilation factor of 4 to handle huge historical information. Hence, the TCN-I network can analyze the hidden long-term dependencies within the dataset to get forecasting results. The model can effectively manage the multivariate dataset, handle the missing values, and perform forecasting tasks effectively.

Research experiments revealed that incorporating various meteorological factors significantly enhances PM2.5 forecasting accuracy by 84%. This highlights the substantial impact of these variables on prediction outcomes. Additionally, our proposed TCN-I model, which utilizes Temporal Convolutional Network (TCN) technology for data imputation, demonstrated a remarkable 78% improvement in forecasting accuracy compared to traditional mean imputation methods.

In the future, the model can be extended to multiple sites to analyze the interaction among different air pollution monitoring stations and improve the forecasting results. Further, the model performs a site-specific p rediction t ask. S o u tilizing t he m odel for regional forecasting tasks can be more useful in the future.

References

[1] Ma, J., Ding, Y., Cheng, J.C., Jiang, F., Gan, V.J. and Xu, Z. (2020) A lag-flstm deep learning network based on bayesian optimization for multi-sequential-variant pm2. 5 prediction. *Sustainable Cities and Society* **60**: 102237.

[2] Zíková, N., Wang, Y., Yang, F., Li, X., Tian, M. and Hopke, P.K. (2016) On the source contribution to beijing pm2. 5 concentrations. *Atmospheric Environment* **134**: 84–95.

[3] Samal, K.K.R., Panda, A.K., Babu, K.S. and Das, S.K. (2021) An improved pollution forecasting model with meteorological impact using multiple imputation and fine-tuning approach. *Sustainable Cities and Society* : 102923.

[4] Samal, K.K.R., Babu, K.S. and Das, S.K. (2020) Ors: The optimal routing solution for smart city users. In *Electronic Systems and Intelligent Computing* (Springer), 177–186.

[5] Yang, J., Shi, B., Shi, Y., Marvin, S., Zheng, Y. and Xia, G. (2020) Air pollution dispersal in high density urban areas: Research on the triadic relation of wind, air pollution, and urban form. *Sustainable Cities and Society* **54**: 101941.

[6] Reiminger, N., Jurado, X., Vazquez, J., Wemmert, C., Blond, N., Wertel, J. and Dufresne, M. (2020) Methodologies to assess mean annual air pollution concentration combining numerical results and wind roses. *Sustainable Cities and Society* : 102221.

[7] Samal, K.K.R., Babu, K.S. and Das, S.K. (2021) Temporal convolutional denoising autoencoder network for air pollution prediction with missing values. *Urban Climate* **38**: 100872.

[8] Zhang, R., Liu, C., Hsu, P.C., Zhang, C., Liu, N., Zhang, J., Lee, H.R. *et al.* (2016) Nanofiber air filters with high-temperature stability for efficient pm2. 5 removal from the pollution sources. *Nano letters* **16**(6): 3642–3649.

[9] Kalisa, E., Fadlallah, S., Amani, M., Nahayo, L. and Habiyaremye, G. (2018) Temperature and air pollution relationship during heatwaves in birmingham, uk. *Sustainable cities and society* **43**: 111–120.

[10] Askariyeh, M.H., Zietsman, J. and Autenrieth, R. (2020) Traffic contribution to pm2. 5 increment in the near-road environment. *Atmospheric Environment* **224**: 117113.

[11] Hien, P., Bac, V., Tham, H., Nhan, D. and Vinh, L. (2002) Influence of meteorological conditions on pm2. 5 and pm2. 5- 10 concentrations during the monsoon season in hanoi, vietnam. *Atmospheric Environment* **36**(21): 3473–3484.

[12] Cheng, Y., He, K.b., Du, Z.y., Zheng, M., Duan, F.k. and Ma, Y.l. (2015) Humidity plays an important role in the pm2. 5 pollution in beijing. *Environmental pollution* **197**: 68–75.

[13] Zalakeviciute, R., López-Villada, J. and Rybarczyk, Y. (2018) Contrasted effects of relative humidity and precipitation on urban pm2. 5 pollution in high elevation urban areas. *Sustainability* **10**(6): 2064.

[14] Amarpuri, L., Yadav, N., Kumar, G. and Agrawal, S. (2019) Prediction of co 2 emissions using deep learning hybrid approach: A case study in indian context. In *2019 Twelfth International Conference on Contemporary Computing (IC3)* (IEEE): 1–6.

[15] Xu, X. and Yoneda, M. (2019) Multitask air-quality prediction based on lstm-autoencoder model. *IEEE transactions on cybernetics* .

[16] Samal, K.K.R., Babu, K.S. and Das, S.K. (2021) Multi-directional temporal convolutional artificial neural network for pm2. 5 forecasting with missing values: A deep learning approach. *Urban Climate* **36**: 100800.

[17] Samal, K.K.R., Babu, K.S., Das, S.K. and 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*: 80–85.

[18] Lagesse, B., Wang, S., Larson, T.V. and Kim, A.A. (2020) Predicting pm2. 5 in well-mixed indoor air for a large office building using regression and artificial neural network models. *Environmental Science & Technology* **54**(23): 15320–15328.

[19] Freeman, B.S., Taylor, G., Gharabaghi, B. and Thé, J. (2018) Forecasting air quality time series using deep learning. *Journal of the Air & Waste Management Association* **68**(8): 866–886.

[20] Samal, K., Babu, K.S. and Das, S.K. (2021) Spatio-temporal prediction of air quality using distance based interpolation and deep learning techniques. *EAI Endorsed Transactions on Smart Cities* **5**(14): e4.

[21] Samal, K.K.R., Babu, K.S., Acharya, A. and Das, S.K. (2020) Long term forecasting of ambient air quality using deep learning approach. In *2020 IEEE 17th India Council International Conference (INDICON)* (IEEE): 1–6.

[22] Ge, L., Zhou, A., Li, H. and Liu, J. (2019) Spatially fine-grained air quality prediction based on dbu-lstm. In *Proceedings of the 16th ACM International Conference on Computing Frontiers* (ACM): 202–205.

[23] Yeo, I., Choi, Y., Lops, Y. and Sayeed, A. Efficient pm2. 5 forecasting using geographical correlation based on integrated deep learning algorithms .

[24] Chen, H., Guan, M. and Li, H. (2021) Air quality prediction based on integrated dual lstm model. *IEEE Access* **9**: 93285–93297.

[25] Xie, H., Ji, L., Wang, Q. and 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), **300**: 032073.

[26] Gilik, A., Ogrenci, A.S. and Ozmen, A. (2021) Air quality prediction using cnn+ lstm-based hybrid deep learning architecture. *Environmental Science and Pollution Research* : 1–19.

[27] Samal, K.K.R.S., Babu, K.S. and Das, S.K. (2023) Spatial-temporal prediction of air quality by deep learning and kriging interpolation approach. *EAI Endorsed Transactions on Scalable Information Systems* **10**(5).

[28] Quinteros, M.E., Lu, S., Blazquez, C., Cárdenas-R, J.P., Ossa, X., Delgado-Saborit, J.M., Harrison, R.M. *et al.* (2019) Use of data imputation tools to reconstruct incomplete air quality datasets: A case-study in temuco, chile. *Atmospheric environment* **200**: 40–49.

[29] Peña, M., Ortega, P. and Orellana, M. (2019) A novel imputation method for missing values in air pollutant time series data. In *2019 IEEE Latin American Conference on Computational Intelligence (LA-CCI)* (IEEE): 1–6.

[30] Yen, N.Y., Chang, J.W., Liao, J.Y. and Yong, Y.M. (2019) Analysis of interpolation algorithms for the missing values in iot time series: a case of air quality in taiwan. *The Journal of Supercomputing* : 1–26.

[31] Wijesekara, W. and Liyanage, L. (2020) Comparison of imputation methods for missing values in air pollution data: Case study on sydney air quality index. In *Future of Information and Communication Conference* (Springer): 257–269.

[32] Libasin, Z., Ul-Saufie, A.Z., Ahmat, H. and Shaziayani, W.N. (2020) Single and multiple imputation method to replace missing values in air pollution datasets: A review. In *IOP Conference Series: Earth and Environmental Science* (IOP Publishing), **616**: 012002.

[33] Lelieveld, J., Evans, J.S., Fnais, M., Giannadaki, D. and Pozzer, A. (2015) The contribution of outdoor air pollution sources to premature mortality on a global scale. *Nature* **525**(7569): 367–371.

[34] Zhai, B. and Chen, J. (2018) Development of a stacked ensemble model for forecasting and analyzing daily average pm2. 5 concentrations in beijing, china. *Science of The Total Environment* **635**: 644–658.
[35] Zhang, S., Guo, B., Dong, A., He, J., Xu, Z. and Chen, S.X. (2017) Cautionary tales on air-quality improvement in beijing. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **473**(2205): 20170457.
[36] Chen, S. (2017), Beijing multi-site air-quality data data set.
[37] Samal, K.K.R., Babu, K.S., Panda, A.K. and Das, S.K. (2020) Data driven multivariate air quality forecasting using dynamic fine tuning autoencoder layer. In *2020 IEEE 17th India Council International Conference (INDICON)* (IEEE): 1–6.
[38] Stekhoven, D.J. and Bühlmann, P. (2012) Missforest—non-parametric missing value imputation for mixed-type data. *Bioinformatics* **28**(1): 112–118.
[39] Malarvizhi, M.R. and Thanamani, A.S. (2012) K-nearest neighbor in missing data imputation. *International Journal of Engineering Research and Development* **5**(1): 5–7.
[40] Beretta, L. and Santaniello, A. (2016) Nearest neighbor imputation algorithms: a critical evaluation. *BMC medical informatics and decision making* **16**(3): 74.
[41] Mustafa, Y.T., Tolpekin, V.A. and Stein, A. (2011) Application of the expectation maximization algorithm to estimate missing values in gaussian bayesian network modeling for forest growth. *IEEE transactions on geoscience and remote sensing* **50**(5): 1821–1831.
[42] Che, Z., Purushotham, S., Cho, K., Sontag, D. and Liu, Y. (2018) Recurrent neural networks for multivariate time series with missing values. *Scientific reports* **8**(1): 1–12.
[43] Gondara, L. and Wang, K. (2018) Mida: Multiple imputation using denoising autoencoders. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining* (Springer): 260–272.
[44] Singh, D. and Singh, B. (2020) Investigating the impact of data normalization on classification performance. *Applied Soft Computing* **97**: 105524.
[45] Guo, G. and Yuan, W. (2020) Short-term traffic speed forecasting based on graph attention temporal convolutional networks. *Neurocomputing* **410**: 387–393.
[46] Pandey, A. and Wang, D. (2019) Tcnn: Temporal convolutional neural network for real-time speech enhancement in the time domain. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (IEEE): 6875–6879.
[47] Meng, C., Jiang, X.S., Wei, X.M. and Wei, T. (2020) A time convolutional network based outlier detection for multidimensional time series in cyber-physical-social systems. *IEEE Access* **8**: 74933–74942.
[48] Bai, S., Kolter, J.Z. and Koltun, V. (2018) An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint arXiv:1803.01271* .
[49] Liu, H. and Chen, C. (2020) Spatial air quality index prediction model based on decomposition, adaptive boosting, and three-stage feature selection: A case study in china. *Journal of Cleaner Production* : 121777.
[50] Qiao, W., Moayedi, H. and Foong, L.K. (2020) Nature-inspired hybrid techniques of iwo, da, es, ga, and ica, validated through a k-fold validation process predicting monthly natural gas consumption. *Energy and Buildings* : 110023.
[51] Ausati, S. and Amanollahi, J. (2016) Assessing the accuracy of anfis, eemd-grnn, pcr, and mlr models in predicting pm2. 5. *Atmospheric environment* **142**: 465–474.
[52] Du, S., Li, T., Yang, Y. and Horng, S.J. (2018) Deep air quality forecasting using hybrid deep learning framework. *arXiv preprint arXiv:1812.04783* .
[53] Ma, J., Ding, Y., Cheng, J.C., Jiang, F. and Wan, Z. (2019) A temporal-spatial interpolation and extrapolation method based on geographic long short-term memory neural network for pm2. 5. *Journal of Cleaner Production* **237**: 117729.
[54] Zhang, B., Zhang, H., Zhao, G. and Lian, J. (2020) Constructing a pm2. 5 concentration prediction model by combining auto-encoder with bi-lstm neural networks. *Environmental Modelling & Software* **124**: 104600.
[55] Tao, Q., Liu, F., Li, Y. and Sidorov, D. (2019) Air pollution forecasting using a deep learning model based on 1d convnets and bidirectional gru. *IEEE Access* **7**: 76690–76698.