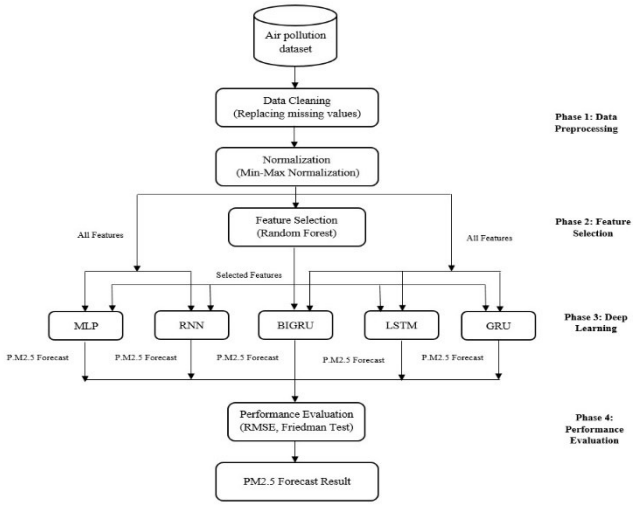
Global NEST Journal, Vol 26, No 3, 05631

Prediction of particulate matter PM2.5 using bidirectional gated recurrent unit with feature selection

Siva Sankari Subbiah[1], Senthil Kumar Paramasivan[2,*] and Muthamilselvan Thangavel[2]
[1]School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, 632014, Tamil Nadu, India
[2]School of Computer Science Engineering and Information Systems, Vellore Institute of Technology, Vellore, 632014,Tamil Nadu, India
Received: 03/12/2023, Accepted: 01/01/2024, Available online: 13/01/2024
*to whom all correspondence should be addressed: e-mail: senthilkumarpresearch@gmail.com
https://doi.org/10.30955/gnj.005631

Graphical abstracts



Abstract

In recent years, air pollution has increased with industrialization and urbanization globally. It is an important hazardous factor that causes severe health issues to community's health. Among the number of pollutants in air, PM2.5 is very dangerous due to its very small, 2.5µm, diameter. The PM2.5 concentration in air causes severe life-threatening to humans. In this paper, RFBIGRU model is proposed to predict PM2.5 in the atmospheric air. RFBIGRU improves PM2.5 prediction accuracy using Random Forest (RF) feature selector and Bidirectional Gated Recurrent Unit (BIGRU) deep neural network. The PM2.5 concentration in air depends on other pollutants' concentration in the air. However, the consideration of several other pollutants increases the curse of dimensionality and overfitting issues. So, in RFBIGRU, first, the relevant pollutants to PM2.5 are identified using random forest feature importance. Then the nonlinear and temporal patterns of the time series air pollutant data are extracted both in forward and backward direction using Bidirectional GRU. The RFBIGRU reduces the curse of dimensionality, overfitting and improves the PM2.5 prediction accuracy compared to other deep learning methods. The experimental result proves RFBIGRU outperforms others by producing least Root Mean Square Error of 42.217 and 6.813 for Delhi and Amaravathi regions.

Keywords: Air pollution, air quality index, deep learning, machine learning, random forest

1. Introduction

In modern era, the rapid development of technology, industry, population and the urbanization of cities raises air pollution. The global ecosystem is also affected by the emission of greenhouse gases due to anthropogenic activities. The amount of air pollutants increases every day and causes major environmental issues. It affects humans indirectly by penetrating through agricultural food products. It causes discomfort for the living organisms and citizens (Alegria *et al.*, 1991, Ercilla-Montserrat *et al.*, 2018). It threatens human being health and the country's economy very worse. The air quality depends on many factors such as weather, every day behavior of human beings and the usage of land in the surroundings. It creates the situation for the people to breathe the worst polluted air every day. Hence, the metropolitan development pattern is assessed that 70% of the total populace will live in metropolitan urban areas in 2050 (Obando Bobadilla *et al.*, 2018). The growth of urbanization increases the air pollution that becomes a major challenge for the growth of developing countries like India and China. Many steps are taken by governments and private organizations to establish air quality monitoring stations. But, still it requires some scientific predictions to control air pollution, to protect human beings by initiating an early warning system and planning outdoor activities (Ma *et al.*, 2019).

The environmental air comprises of a different type of poisons as pollutants like particulate matter PM2.5, particulate matter PM10, sulfur dioxide (SO_2), Ozone (O_3), carbon monoxide (CO) and Nitrogen dioxide (NO_2). The concentration of these pollutants has a direct relationship with the healthiness of human beings. The level of air pollution increases in the atmospheric air when there is an increase of the concentration of the pollutants level in the air which in turn increases the severity of the human

beings' health issues (Ding *et al.*, 2022, Aliyu *et al.*, 2014). Among all pollutants, particulate matter (PM) creates a worse effect on human beings when they are exposed to contaminated air. The major origins of the particulate matter are industries, wood-burning heaters and smoke from motor vehicles. The PM consists of a variety of components like organic chemicals, dust particles, metals, sulfates and nitrates. The concentration of PM increases when dust storm and bushfires occur.

There are two types of particulate matter namely, PM10 and PM2.5 based on the size of the particle. The size of PM10 is 10 micrometers in diameter whereas the PM2.5 size is about 2.5 micrometers in diameter. PM10 can enter into lungs through the nose and throat due to its small size. PM2.5 is so small compared to PM10. It can enter into bloodstream through the nose, throat and lungs. It causes adverse health problems for humans. The increase in the concentration of particulate matter increases heart diseases and lung diseases which increases the ratio of human death. Even if a human is exposed to the particulate matter for a short duration, it creates severe irritations in the nose, throat and eye, worsens lung diseases, triggers heart attacks and increases per-mature deaths. For the long term exposure, reduces lung functionality, develops cardiovascular diseases and reduces life expectancy. So, it mandates the necessity of forecasting the particulate matter (Aliyu *et al.*, 2014). The nomenclature utilized in this paper is given in Table 1.

PM2.5 can be predicted using statisticall models, machine learning and deep learning models. Initially, the statisticall models depended on numerical prediction, statistics and probability. They are more complicated and less efficient. Next, artificial intelligence based machine learning models like artificial neural network (ANN), radial basis function (RBF), Multilayer perceptron (MLP), support vector machine (SVM), random forest (RF), linear regression (LR) and multiple linear regression (MLR) were utilized. The machine learning-based PM2.5 prediction is more efficient and reliable. The consideration of more environmental factors guarantees an improved prediction of PM2.5. The advent of deep learning models greatly helps to improve PM2.5 air pollutant forecasting by handling the time series nonlinear air pollutant data with numerous other factors related to air pollution. Now a days, many researchers in the literature utilized deep learning-based recurrent neural networks and their variants. They handled incompleteness, uncertainty and non-stationary issues of the time series air pollutant data. The researchers also employed the benefits of utilizing a bidirectional network for improving accuracy. The accuracy can also be increased by considering many features as input (Kiruthika *et al.*, 2014, Swaroop *et al.*, 2014). When the number of features increases the curse of dimensionality issue also increases. It will result in a serious overfitting issue.

In this paper, PM2.5 is forecasted using the history of PM2.5 and other pollutants. Among all pollutants, the relevant pollutants for PM2.5 are identified by using random forest for reducing the overfitting and curse of dimensionality issues. The non-linearity, uncertainty and sequence dependency characteristics remains in the time series PM2.5 pollutant is analyzed. Then the similar patterns which is hidden in the historical pollutant data are extracted for improving the accuracy of PM2.5 prediction using bidirectional gated recurrent unit.

Table 1. Nomenclature

WHO	-	World Health Organization	NO2	-	Nitrogen Dioxide
ADS	-	Amaravathi Region Data Set	NOx	-	Nitrogen Oxides
AI	-	Artificial Intelligence	O3	-	Ozone
ANN	-	Artificial Neural Network	PM10	-	Particulate Matter PM10
AQI	-	Air Quality Index	PM2.5	-	Particulate Matter PM2.5
ARIMA	-	Autoregressive Integrated Moving Average	RBF	-	Radial Basis Function
			RF	-	Random Forest
ARMA	-	Autoregressive Moving Average	RFBIGRU	-	BIdirectional Gated Recurrent Unit with Random Forest
Bi-LSTM	-	Bidirectional Long Short Term Memory	RFGRU	-	Gated Recurrent Unit with Random Forest
CO	-	Carbon Monoxide	RFLSTM	-	Long Short Term Memory with Random Forest
DDS	-	Delhi Region Data Set	RFMLP	-	Multilayer Perceptron with Random Forest
DL	-	Deep Learning	RFRNN	-	Recurrent Neural Network with Random Forest
EMD	-	Empirical Mode Decomposition	RMSE	-	Root Mean Square Error
GRA	-	Grey Relational Analysis	RNN	-	Recurrent Neural Network
GRU	-	Gated Recurrent Unit	SMOTE	-	Synthetic Minority Oversampling Technique
LR	-	Linear Regression	SO2	-	Sulfur Dioxide
LSTM	-	Long Short Term Memory	SVM	-	Support Vector Machine
MAE	-	Mean Absolute Error	SVR	-	Support Vector Regression
ML	-	Machine Learning			
MLP	-	Multilayer Perceptron			
MLR	-	Multiple Linear Regression			
NH3	-	Ammonia			
NO	-	Nitric Oxide			

The contribution of the paper is organized as follows,

- *Data Preprocessing (Remove incompleteness):* To improve the input air pollutant data quality by finding missing values and filling it with mean values.
- *Feature Selection (Eliminate Curse of Dimensionality)*: To select the relevant features by employing a random forest model and finding the feature importance.
- *Deep Learning (Improve accuracy)*: To improve accuracy by handling temporal dependency at the next and previous steps using bidirectional gated recurrent unit
- *Performance Evaluation (Proving Superiority)*: To prove the superiority of the proposed RFBIGRU model against the regression-based state-of-the-art deep learning models such as MLP, RNN, LSTM and GRU in terms of RMSE and Friedman test

The rest of the present paper is structured as follows. Section 2 reviews the related works done and finds the research gap in air pollution prediction. Section 3 discusses the proposed RFBIGRU architecture and describes the competing methodologies. Section 4 presents the experimental results using Delhi and Amaravathi air pollutants datasets. Section 5 concludes the present research.

2. Related Works

This section discusses the research work that has been done by various researchers in the field of air quality prediction. It is categorized into the research done using statistical and machine learning techniques. The widely used statistical methods are Autoregressive Integrated Moving Average (ARIMA) (Kumar and Jain 2010, Bedekar *et al.*, 2021), Autoregressive Moving Average (ARMA) (Box *et al.*, 2015) and multiple linear regression (MLR) (Li *et al.*, 2011). The simple linear regression (LR) was the widely used statistical model in earlier days for predicting the

PM2.5 concentration. Mani and Viswanadhapalli (2022) designed an air quality prediction model using two time series models namely multiple linear regression and ARIMA. The author predicted the air quality index (AQI) for the next 15 days using the data collected from Chennai, India. The result shows that MLR achieved 92% and ARIMA achieved 95% accurate prediction of air quality. Amelia *et al.* (2022) developed a PM2.5 prediction model using seasonal ARIMA and triple exponential smoothing time series models. The experimental result shows the seasonal ARIMA produced better prediction than exponential smoothing. It can handle the data with a linear relationship between dependent and independent attributes. But in reality, the concentration of PM2.5 does not have a linear nature. Mostly it is an irregular sequence and non-linear nature. The mixed logic regression model, ARMA, ARIMA and MLR produce an accurate forecast only when the concentration of fine particles is in a linear nature. So these statistical models fail to produce an accurate prediction of PM2.5 with large datasets of a nonlinear nature (Kumar and Pande 2022, Xu *et al.*, 2022).

After the advent of artificial intelligence (AI) techniques, many researchers have contributed by designing a variety of machine learning (ML) models of single and hybrid category to handle nonlinearity of air pollutant data and improve the accuracy of PM2.5 prediction (Zhan *et al.* 2017, Lu *et al.*, 2021, Li *et al.*, 2021, Xiao *et al.*, 2018). The artificial neural network (ANN), support vector regression (SVR) and random forest (RF) are the widely utilized methodologies to predict PM2.5 concentration in the air (Yu *et al.*, 2016, Lin *et al.*, 2011, Wang *et al.*, 2015). Saiohai *et al.* (2023) presented a study and compared the prediction of PM2.5 using multilayer perceptron (MLP) and multiple linear regression (MLR). Masood and Ahmad (2020) presented a model for predicting PM2.5 concentration daily using two machine learning models namely SVM and ANN. The two years of air pollutants and meteorological data were utilized for the experimental analysis. The results produced by ANN show better improvement than SVM. Even though the ML models handle the nonlinearity nature of air pollutant data, it is not free from some limitations like overfitting and local optimization. When the dataset size and attributes count in the dataset increases it will increase the network complexity, overfitting and the gradient decent problems.

In recent years, the type of machine learning called deep learning (DL) models invented by researchers to overcome such an issue of machine learning (Siva Sankari and Senthil Kumar 2023). It is a large artificial neural network with multiple hidden layers that can handle complex problems like nonlinearity, uncertainty and overfitting (Senthil Kumar 2019). The historical air pollutant data can have a similar air pollutant concentration pattern in the future (Pruthi and Liu 2022). So it has a sequence dependence nature that can be effectively handled by the type of deep neural network called RNN and its variants called LSTM and GRU. Kim *et al.*, (2023) presented a model to forecast PM2.5 for the next one hour using deep learning-based bidirectional LSTM with random forest. The author handled the imbalanced data well and produced an improved PM2.5 forecast than the state-of-the-art ML and DL methods. Huang *et al.* (2021) designed a PM2.5 prediction model using empirical mode decomposition (EMD) and GRU. The author considered the non-stationary characteristics of the air pollutants data and the meteorological data by decomposing the sequence using empirical mode decomposition (EMD) and then fed each decomposed sequence to the GRU along with the meteorological data for training and forecasting. The result shows the EMD-GRU reduced the forecasting error compared to a single GRU. Qing (2023) developed a model for forecasting PM2.5 using grey relational analysis (GRA) and GRU. First, the meteorological features are compared against PM2.5. Then the reference sequence was constructed from the PM2.5 of the monitoring stations and grey correlation analysis was employed to compare and construct a spatial weight matrix. Followed by, the weight matrix was utilized for extracting the spatial relationship. After that, the forecasting was performed using GRU. The result showed that the GRU achieved better accuracy in reduced time compared to LSTM.

Most of the researchers in the literature utilized deep learning-based recurrent neural networks and their variants. They handled the incompleteness and non-stationary issues of the time series air pollutant data. The researchers also employed the benefits of utilizing a bidirectional network for improving accuracy. The accuracy can also be increased by considering many features as input. The increase of input features will result the curse of dimensionality and a serious overfitting issues in the prediction process. Hence, the consideration of irrelevant features may slow down the learning process and prediction accuracy. So, identification of the relevant features is important to improve the accuracy by preventing from misguiding during the learning process. Similarly, instead of GRU, the considerations of Bidirectional GRU improves the accuracy by processing the input data in both forward & backward directions and utilizing both the previous and next computational information at every time step. In this paper, the issues such as overfitting and curse of dimensionality are addressed by identifying the relevant air pollutant features of PM2.5 using random forest. The accuracy of the pm2.5 prediction is improved by handling nonlinearity, uncertainty and sequence dependency in the air pollutant data using bidirectional gated recurrent unit.

3. RFBIGRU model

The objective of designing the RFBIGRU model is to predict the PM2.5 pollutant accurately. It comprises of data preprocessing, feature selection, deep learning and performance evaluation phases. The following sections discuss the significance of the proposed RFBIGRU model and its architecture.

3.1. Significance of RFBIGRU

The RFBIGRU is a hybrid model which combines feature selection and deep learning concepts for an accurate prediction of air pollutant PM2.5. It helps to handle issues

such as overfitting, curse of dimensionality, non-linearity and uncertainty during the prediction process. The real time air pollutant data may have some incompleteness. In addition to that the pollutant data is nonlinear and uncertain in nature. Hence, the number of features to be utilized as the determinant features for PM2.5 also plays a role in the accuracy of the prediction process. The increase of the number of supporting features will increase the accuracy of the prediction also. But it will create a chance for the curse of dimensionality. Due to the curse of dimensionality, the consideration of less important features may take down the accuracy and increase the complexity. In this paper, the RFBIGRU model performs data cleaning to remove the incompleteness and min-max normalization to transform all data to the unified range which is suitable for prediction process. Followed by the powerful task of identifying the relevant features to PM2.5 is carried out using random forest to eliminate the curse of dimensionality and reduce the overfitting issues. Finally, it handles the non-linearity, uncertainty of pollutants data and finds the patterns hidden inside using the dominant deep learning methodology, bidirectional gated recurrent unit.

3.2. RFBIGRU architecture

The primary goal of RFBIGRU is to predict PM2.5 which is composed of four phases. The first phase is the data preprocessing phase. In this phase, the preliminary processes to enhance the quality of the input data are carried out. The missing values are identified and filled with the mean values. Followed by all the pollutant data are transformed using min-max normalization. Next phase is the feature selection phase. In this phase, the model complexity is reduced by selecting relevant features for the prediction process. All the features do not guarantee to improve the prediction accuracy. The minimal features that provide an efficient contribution to improving the prediction accuracy are identified before starting the prediction process. So the relevant features are selected by designing the random forest model. The feature importance calculated by the random forest feature selector is considered as the weight of those features towards the PM2.5. The features with high importance are selected as the relevant features.

The third phase is the deep learning phase which performs the actual prediction of PM 2.5 pollutant in the atmospheric air. The recorded pollutant data is time series in nature. The temporal dependency in the recorded data retains the future trend of the pollutant data. Hence the uncertain and non-linearity of the data should also be analyzed properly for an accurate prediction of PM2.5. So, the deep learning-based bidirectional GRU is employed for the prediction of PM2.5. The regression-based deep learning methods such as MLP, RNN, LSTM and GRU are utilized as the competing models for BIGRU. The performance of these all models also can be improved by providing the required relevant features using random forest. The fourth phase is the performance evaluation phase. It measures the prediction performance of all methodologies and compares RFBIGRU performance against all other methodologies with and without random forest selected features in terms of RMSE and conducting the Friedman test. The architecture of RFBIGRU is given in Figure 1.

3.2.1. Data Preprocessing

The outcome of the prediction model depends on the input provided to the model. So, the accurate prediction results can be attained by providing the high quality data as an input. The air pollutant data may have incompleteness which will misguide the prediction process. So, the incomplete missing values in the dataset are replaced by the mean values of the past week's data. Then the data is transformed to a unified format by applying min-max normalization. It will transform all the data in the dataset to the range 0 to 1. Let 'f' denotes the feature, 'minimum_f' denotes the minimum of feature 'f' and 'maximum_f' denotes the maximum of feature 'f'. The actual interval is denoted as [minimum_f, maximum_f] and the transformed new interval is denoted as [newMin_f, newMAX_f]. The data 'd' from the actual interval is transformed and mapped to the newly transformed data 'newData_d' in the normalized interval [0, 1]. It is defined as follows,

$$\text{newData}_d = \frac{d - \text{minimum}_f}{\text{maximum}_f - \text{minimum}_f} \left(\text{newMAX}_f - \text{newMin}_f\right) + \text{newMin}_f \quad (1)$$

After normalization, the input air pollutant data is suitable for further feature selection and prediction processes.

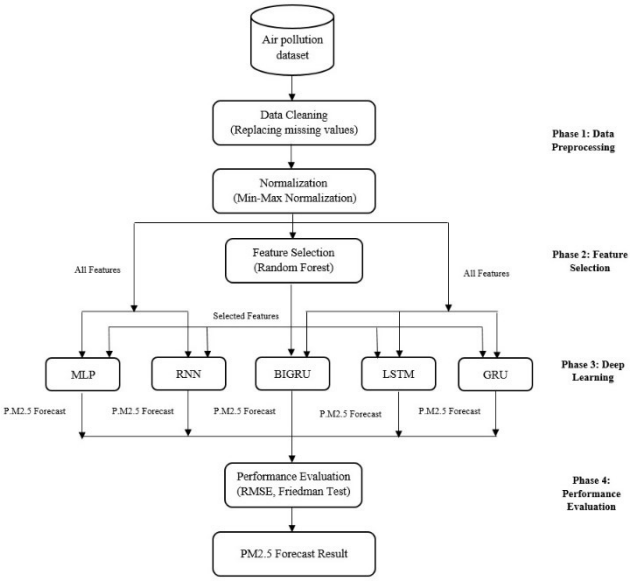


Figure 1. RFBIGRU Architecture

3.2.2. Feature Selection

The feature selection is an important dimensionality reduction technique. In this paper, the dimension of the dataset is reduced vertically. It identifies less important and irrelevant features and removing those features from the input dataset. Hence the complexity of the prediction process is reduced and the time taken for the computation also reduces with reduced input features (Subbiah and Chinnappan 2021). In this paper, the

reduction of dataset dimension is done by finding the less important features using random forest. It calculates the importance of each feature as node purity. In regression problems, the random forest calculates the MSE as the node purity. The increase in node purity represents the increase in the importance of the feature. So, the features with less node purity are identified as less important features and are not considered as input to the prediction process (Chen *et al.*, 2020).

Random forest: The random forest is one of the embedded categories of feature selection. It combines model-based (wrapper) and model-free (filter) feature selection concepts together. The model-free method finds the relevant features by calculating the weight of each feature using its inherent statistical characteristics. The weight is higher the relevance and importance of that feature are also higher. The model-based feature selection finds the subset of features that provides the highest model accuracy using the model as the internal feature selector. The subset with minimal error or highest accuracy is considered as the selected features (Subbiah *et al.*, 2023)]. Thus, the random forest has both characteristics of feature selection. Random forest is the supervised category of machine learning model which follows the decision tree and bagging concepts. It builds the bag of samples by taking random samples and random features from the input dataset. Then it builds a decision tree using these bags. It calculates the importance of each feature and then builds the node for the decision tree. Likewise, for all the bags, it constructs decision trees. Normally, the tree with the topmost nodes has higher importance than the nodes at the bottom levels of the tree. Finally, it calculates the importance by measuring the average of all the trees constructed. As the random forest follows the bagging it is less prone to overfitting and contributes much to improving the prediction accuracy (Guo *et al.*, 2023).

3.2.3. Deep learning

Deep learning methods can handle non-linear, seasonal, sequence-dependent air pollutant data effectively (Drewil and Al-Bahadili 2022, Subbiah and Chinnappan 2020). The time series air pollutant data has the internal sequence dependency behavior between different pollutants. Similar patterns of air pollutant data in history may appear in the future. It will be utilized to predict the air pollutant levels (Bouktif *et al.*, 2020, Vijayanti *et al.*, 2023). The following section discusses the deep learning methodologies such as MLP, RNN, LSTM, GRU and BIGRU employed for predicting the air pollutant PM2.5.

Multilayer perceptron: MLP creates complex non-linear models which is suitable for predicting the air quality. It simulates the functionalities of the human brain. It has one input and one output layer and several hidden layers. Every layer has several neurons to process the data. For each input, there is one neuron at the input layer. Similarly, there is one neuron at the output layer for each output. The data processed in each layer is forwarded to the next layer in the forward direction (Manan *et al.*, 2021). The data provided at the input layer is processed and flows through the hidden layer to the output layer. The error is calculated at the output layer by comparing the generated output against the expected output. Then the calculated error is propagated in backward direction and the weight is adjusted in each layer (Xiaogang and Xin 2022). Let X = (X_1, X_2, X_3,…, X_{N-1}, X_N) be the input at the input layer and 'O' be the output. The H = (H_1, H_2, H_3, … H_{N-1}, H_N) be the hidden vector sequence. The output at the hidden layer 'O_H' is as follows,

$$O = f\left(X\right) \tag{2}$$

$$O_H = \sum_{i=1}^{N} XW_{ih} + W_{ho} \tag{3}$$

$$O = \sum_{i=1}^{k} z_i w_{ho} + w_{l0} \tag{4}$$

Where 'f' represents the activation function that triggers the neurons in the network and 'N' represents the total number of units. The 'W_{ih}' and 'W_{ho}' are the weight matrices. The 'W_{ho}' and 'W_{lo}' are biases at hidden and output layers (Xiaowei 2022).

Recurrent neural network: The RNN is the deep neural network that is designed to process the time series data which has the sequence dependency. The RNN architecture provides the looping facility for looping the hidden state information to maintain the sequence dependency at every computation in hidden unit. Thus, by considering the past input and past computation, it generates the present output at each time step. Unlike MLP, the RNN utilizes the recurrent relations and back propagation through time during forward pass and learning respectively (Zhao *et al.*, 2018). The recurrence relation that updates the hidden state is as follows,

$$h_t = f_w\left(x_t, h_{t-1}\right) \tag{5}$$

After applying the activation function 'h_t' is as follows,

$$h_t = tanh\left(W_{hh}h_{t-1} + W_{xh}x_t\right) \tag{6}$$

The output 'y_t' is as follows

$$y_t = W_{hy}h_t$$

$$y_t = W_{hy}h_t \tag{7}$$

Where 'x' is input, 't' is time step, 'h' is hidden state and 'y' is output.

Long short term memory: The LSTM is the advanced architecture of RNN which can handle long term dependencies that exist in time series data effectively (Ayturan *et al.*, 2018). It overcomes the limitations of RNN network like gradient decent and short term memory capacity. It maintains the longer sequences of hidden state information in each unit for a long period. Hence it is free from gradient issues and keeps longer sequences of hidden state information. It is a powerful logic to analyze and process the sequence dependency. It maintains only the required information and removes the unwanted information in each unit by using three gates namely

forget, input and output (Li *et al.*, 2017, Bekkar *et al.*, 2021). Let 't' be the time, 'x_t' be input at 't', 'h_t' be hidden state at 't' and 'y_t' be output at 't'. The 'f_t', 'i_t' and 'o_t' be forget, input and output gates. The 'ct', 'ct-1' and '$\tilde{c}_t$' be the new cell, previous cell and the candidate. The 'W' and 'b' are weight matrices and biases respectively.

$$f_t = \sigma\left(W_f.\left[h_{t-1}, x_t + b_f\right]\right) \tag{8}$$

$$i_t = \sigma\left(W_i.\left[hta_{t-1}, x_t\right] + b_i\right) \tag{9}$$

$$o_t = \sigma\left(W_o.\left[h_{t-1}, x_t\right] + b_o\right) \tag{10}$$

$$h_t = o_t \odot tanh\left(c_t\right) \tag{11}$$

Where 'c_t' is defined as follows,

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{C}_t \tag{12}$$

Where '$\tilde{C}_t$ is defined as follows,

$$\tilde{C}_t = nh\left(W_c.\left[h_{t-1}, x_t\right] + b_c\right) \tag{13}$$

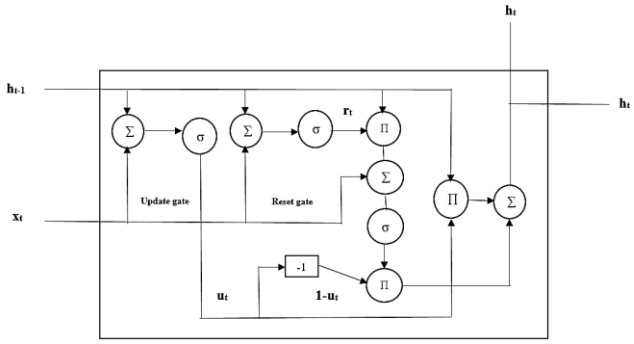


Figure 2. Structure of Gated Recurrent Unit

Gated recurrent unit: The GRU is another extended variant of RNN that also can maintain the hidden state information for a long period and supports the analysis of the time series data (Guo *et al.*, 2023, Subbiah and Senthil Kumar 2022). It can produce better results than LSTM by using the extended architecture in each cell in the network. Instead of three gates as LSTM, it utilizes only two gates namely update and reset, for maintaining the long sequences of hidden state information for a long period. The reset gate finds the unwanted information to be removed and finds the required information from the input. The update gate provides the current input and the previous hidden state information when the sigmoid function is triggered (Athira *et al.*, 2018)]. The architecture of GRU is given in Figure 2. Let 'x' be the input, 'h' be the hidden state, 'u' be the update state, 'r' be the reset state, 'W' be the weight matrix, 'b' be the bias weight matrix and 'σ' be the sigmoid function. The update state at the period 't' is

$$u_t = \sigma\left(W_u x_t + W_u h_{t-1} + b_u\right) \tag{14}$$

$$r_t = \sigma\left(W_r x_t + W_r h_{t-1} + b_r\right) \tag{15}$$

$$h_t = \left(1 - u_t\right) * h_{t-1} + u_t * \tilde{h}_t \tag{16}$$

Where '$\tilde{h}_t$' is defined as follows,

$$\tilde{h}_t = tanh\left(W.\left[r_t * h_{t-1}\right] + W.x_t\right) \tag{17}$$

Bidirectional gated recurrent unit: The BIGRU is a sequence model composed of two GRUs. One GRU processes the input in a forward direction from the beginning of the sequence whereas another one processes in a backward direction from the end of the sequence simultaneously (Xu *et al.*, 2022, Ghose *et al.*, 2022). So, the network can learn the patterns from the previous and subsequent data to process the current input. Figure 3 shows the structure of BiGRU (Liu *et al.*, 2021, Andre Gensler 2019). Let the input be 'x', the hidden state be 'h', the hidden state of the forward GRU at time 't' be '$\overrightarrow{h_t}$' and the hidden state of the backward GRU be '$\overleftarrow{h_t}$'. The BiGRU is defined as follows,

$$\overrightarrow{h_t} = GRU_{forward}\left(x_t, \overrightarrow{h_{t-1}}\right) \tag{18}$$

$$\overleftarrow{h_t} = GRU_{backward}\left(x_t, \overleftarrow{h_{t+1}}\right) \tag{19}$$

$$h_t = \overrightarrow{h_t} \oplus \overleftarrow{h_t} \tag{20}$$

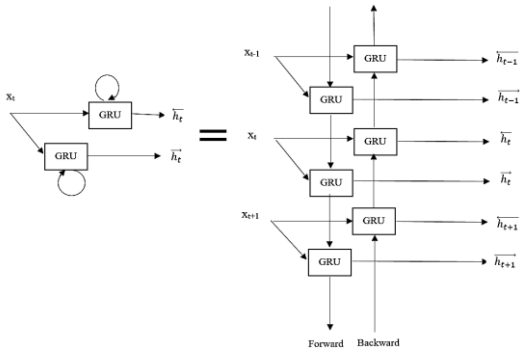


Figure 3. Structure of Bidirectional Gated Recurrent Unit

3.2.4. Performance Evaluation

The performance of the air pollutant PM2.5 prediction can be done in numerous ways. In this paper, the performance of the proposed RFBIGRU model is evaluated by measuring the root mean square error (RMSE). In addition to that the non-parametric significance test namely the Friedman test (multiple comparison test) is also conducted to show the improved performance of RFBIGRU compared to others.

RMSE: The model's goodness can be evaluated by measuring the difference between the actual value and the model's predicted value. Let the total number of samples be denoted as 'N', the actual observed value is denoted as 'actualValue_t' and the forecasted value is denoted as 'forecastValue_t'. The RMSE is defined as follows,

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{t=1}^{N}\left(\text{actualValue}_t - \text{forecastValue}_t\right)^2} \tag{21}$$

Friedman Test: The Friedman test is the nonparametric type of the statistical test. It compares the proposed

RFBIGRU model with all other models. It calculates the difference in forecasting error produced between two or more models. First, it sets the null hypothesis as the same mean value of the error produced by two or more models.

Then the Friedman statistics is calculated as follows,

$$F = \frac{12n}{K(K+1)}\left[\sum_{j=1}^{K} R_j^2 - \frac{K(K+1)^2}{4}\right] \quad (22)$$

where 'F' denotes the Friedman Statistics, 'K' denotes the count of the forecasting models utilized for the comparison, 'n' denotes the number of forecasting values utilized and 'R_j' denotes the average rank sum of the j^{th} forecasting model on the forecasting error 'r'.

$$R_j = \frac{1}{n}\sum_{i=1}^{n} r_i^j \quad (23)$$

Finally, the Friedman statistics is compared against Friedman critical value and the p-value is compared against the significance 'α'. If the Friedman statistics is high and the p-value is less then the null hypothesis is rejected (Ghiasi *et al.*, 2019, Hong 2018, Dong *et al.*, 2018). It shows that the proposed RFBIGRU model outperforms other compared models.

4. Results and discussion

The datasets utilized for the experiment and the forecasting results are presented in this section. It also discusses the results of random forest in finding the relevant features to PM2.5 for improving the forecast performance by reducing the complexity. Hence, it proves the superior performance of proposed RFBIGRU model in terms of root mean square error (RMSE) and Friedman test.

Table 2. Statistical Characteristics of DDS dataset

Feature	Minimum	Maximum	Mean	StdDev
Date	-	-	-	-
PM2.5	10.24	685.36	117.228	82.88
PM10	18.59	796.88	232.863	119.521
NO	3.57	221.03	39.002	33.373
NO2	10.63	162.5	50.8	22.681
NOx	0	254	58	37
NH3	6.78	166.7	42.004	17.264
CO	0	30.44	1.977	2.561
SO2	2.34	71.56	15.904	7.747
O3	6.94	257.73	51.335	25.513
Benzene	0	20.64	3.545	2.432
Toluene	0	103	17.182	15.295
Xylene	0	23.3	1.439	1.859

4.1. Dataset description

The air quality data of two Indian cities Delhi and Amaravathi is considered for the experimental purpose. The air pollutants data of the Delhi region recorded daily from 1st January 2015 to 30th June 2020 and the air pollutant data of Amaravathi from 1st December 2017 to 30th June 2020 are considered for forecasting the particulate matter 2.5 (PM2.5). The datasets consist of 13 features like, date, PM2.5, PM10, NO, NO_2, NO_x, NH_3, CO, SO_2, O_3, Benzene, Toluene and Xylene. The statistical characteristics like minimum, maximum, mean and standard deviation of the features of Delhi region dataset (DDS) and Amaravathi region dataset (ADS) are given in Table 2 and Table 3.

Table 3. Statistical Characteristics of ADS dataset

Feature	Minimum	Maximum	Mean	StdDev
Date	-	-	-	-
PM2.5	4.65	139.38	37.495	27.16
PM10	7.8	230.27	76.068	43.888
NO	0.25	43.76	4.448	4.173
NO2	1.52	140.17	22.174	20.648
NOx	0.86	103.49	15.389	13.132
NH3	0.3	35.2	11.996	5.816
CO	0	1.84	0.632	0.326
SO2	2.9	66.39	14.228	6.789
O3	7.51	137.61	37.489	17.96
Benzene	0	53.89	0.55	3.338
Toluene	0	76.32	1.847	3.517

The visualization of the features of Delhi city air quality data is given in Figure 4. The features of Amaravathi city air quality data are visualized in Figure 5. The dataset is partitioned into training, validation and testing datasets with 70:15:15 instances. First 70% of instances are utilized for training, the next 15% instances are utilized for validation and the next 15% instances are utilized for testing. Totally the dataset has 13 features including the date. The date feature is used as an indexing feature. The PM2.5 is the target feature and the remaining 11 features are determinant features.

4.2. Experimental results

The experiment is conducted by using Python in the Tensorflow environment. All the features in DDS and ADS do not contribute to improving the performance of forecasting the air quality pollutant PM2.5. So, the random forest regression is employed to find the importance of features. It calculates the importance of each feature as node purity. In regression problems, the random forest calculates the MSE as the node purity. The increase in node purity represents the increase in the importance of the feature. The importance of DDS and ADS features in descending order is given in Table 4. It shows the order of high importance to less-importance features of PM2.5 as PM10, Benzene, NO, NH_3, NO_2, CO, SO_2, Toluene, NO_x, O_3 and Xylene.

The importance of features of DDS and ADS are graphically shown in Figure 6. It demonstrates that the feature importance drops steadily from CO feature (5th feature). After that no such major difference in the importance value for the remaining features in both DDS and ADS. So the topmost 5 features PM10, Benzene, NO,

NH_3 and NO_2 are taken into consideration for forecasting PM2.5.

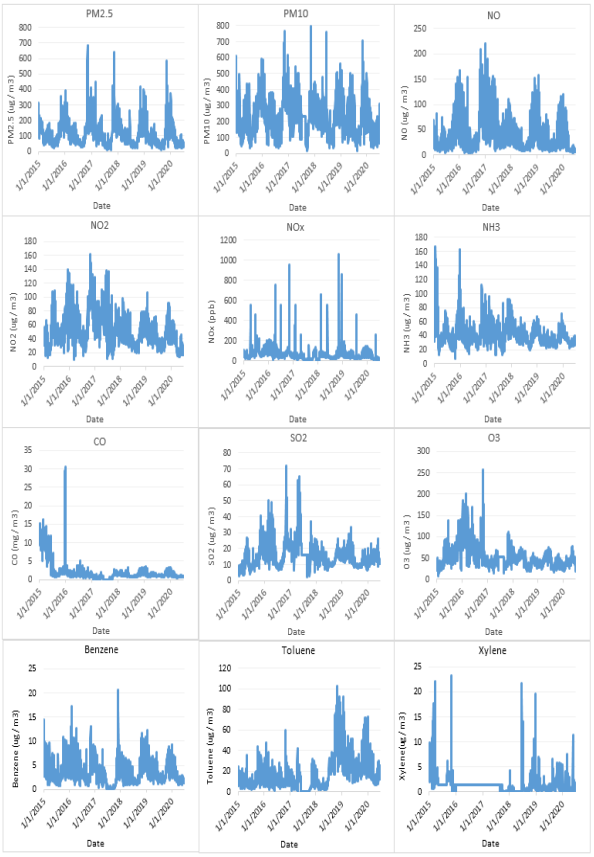


Figure 4. Features of DDS Dataset

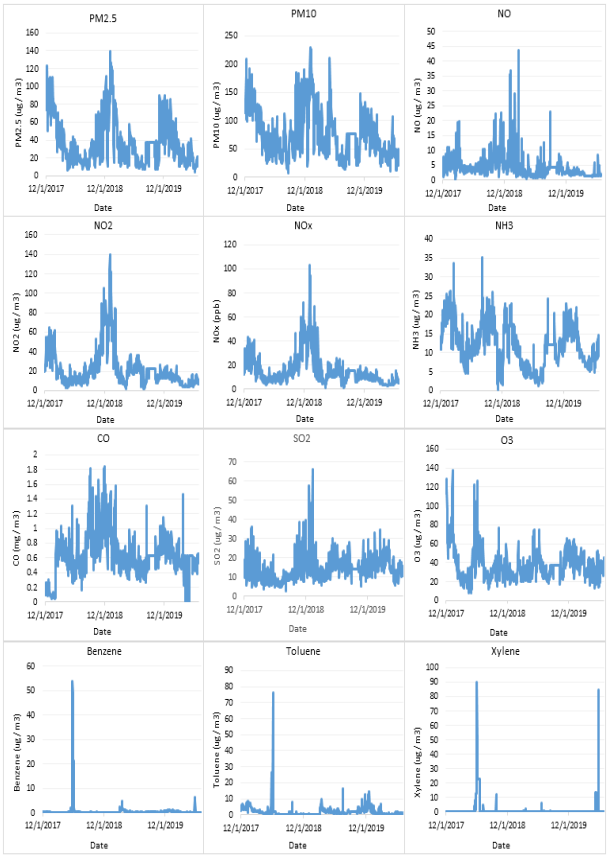


Figure 5. Features of ADS Dataset

Table 4. Importance of DDS and ADS Features

DDS	ADS
PM10	PM10
Benzene	Benzene
NO	NO
NH3	NH3
NO2	NO2
CO	CO
SO2	SO2
Toluene	Toluene
NOx	NOx
O3	O3
Xylene	Xylene

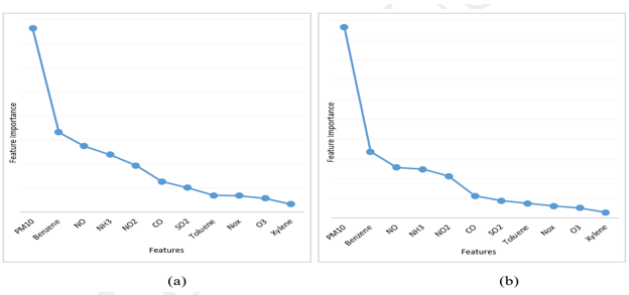


Figure 6. Importance of Features (a) Delhi Dataset (DDS) (b) Amaravathi Dataset (ADS)

The particulate matter PM2.5 is forecasted by designing the bidirectional GRU with selected features using random forest (RFBIGRU). The model is configured as follows. The inputs is 5, the outputs is 1, the hidden layers are 3, the units in the first hidden layer are 30, the units in the second hidden layer are 20 and the units in the third hidden layer are 10, the optimizer is adaptive moment estimation (Adam) and the mean absolute error (MAE) is the loss function. The model is trained and then validated using the time series cross-validation on a rolling basis. RFBIGRU model is tested to forecast PM2.5 using the testing dataset. PM2.5 is also forecasted by using competing models such as MLP, RNN, LSTM, GRU and BIGRU. The forecasting results of MLP, RNN, LSTM, GRU and BIGRU with all features and selected features for DDS are shown in Figure 7.

The PM2.5 forecasting graph of MLP with selected features of random forest in Figure 7(b) has less deviation from actual PM2.5 compared to the forecasting graph of MLP without feature selection in Figure 7(a). Similarly the PM2.5 forecasting graph of RNN, LSTM, GRU and BIGRU with random forest in Figure 7(d), Figure 7(f), Figure 7(h), Figure 7(j) is closer to the actual PM2.5 graph compared to forecasting graph of RNN, LSTM, GRU and BIGRU without feature selection in Figure 7(c), Figure 7(e), Figure 7(g) and Figure 7(i). The example comparison of the PM2.5 forecast results of the proposed RFBIGRU against the actual, MLP, RNN, LSTM, GRU, BIGRU, RFMLP, RFRNN, RFLSTM and RFGRU for DDS is illustrated in Figure 8.

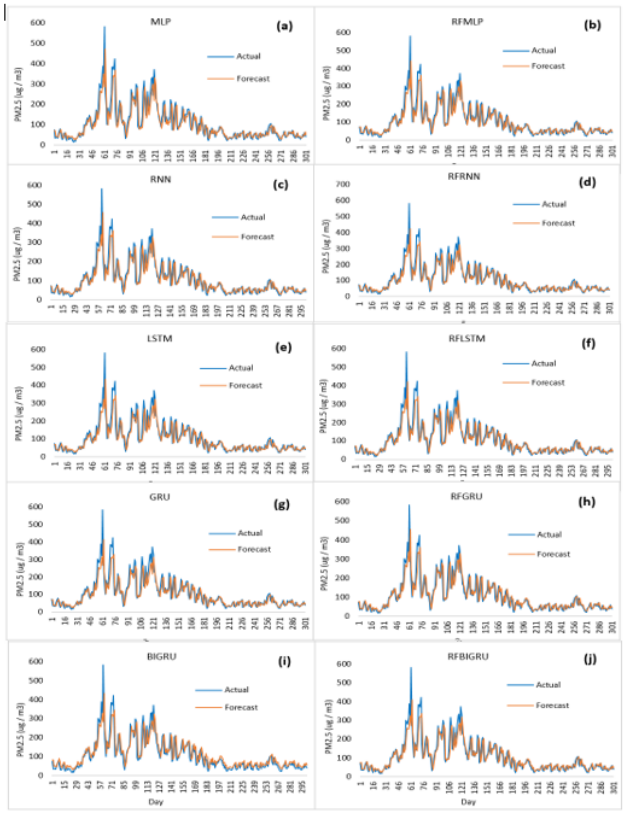


Figure 7. PM2.5 Forecasting Results of DDS without Feature Selection and with Feature Selection

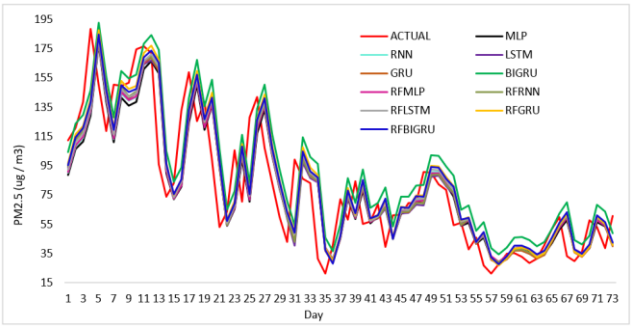


Figure 8. Sample Comparison of PM2.5 Forecasting Results of DDS

Figure 8 proves the improvement of the proposed RFBIGRU PM2.5 forecast result compared to others. In Figure 8, the PM2.5 forecast (blue color) is much closer to actual PM2.5 (red color) compared to all other models. The selected features of the random forest help to improve the bidirectional gated recurrent unit performance in forecasting PM2.5 for DDS. The forecasting results of MLP, RNN, LSTM, GRU and BIGRU with actual features and random forest selected features for ADS are shown in Figure 9. The PM2.5 forecasting graph of MLP with selected features of random forest in Figure 9(b) has less deviation from actual PM2.5 compared to the forecasting graph of MLP without feature selection in Figure 9(a). Similarly, the PM2.5 forecasting graph of RNN, LSTM, GRU and BIGRU with random forest in Figure 9(d), Figure 9(f), Figure 9(h), Figure 9(j) is closer to the actual PM2.5 graph compared to forecasting graph of RNN, LSTM, GRU and BIGRU without feature selection in Figure 9(c), Figure 9(e), Figure 9(g) and Figure 9(i). The example comparison of the PM2.5 forecast results of the proposed RFBIGRU against the actual, MLP, RNN, LSTM, GRU, BIGRU, RFMLP, RFRNN, RFLSTM and RFGRU for ADS is illustrated in Figure 10.

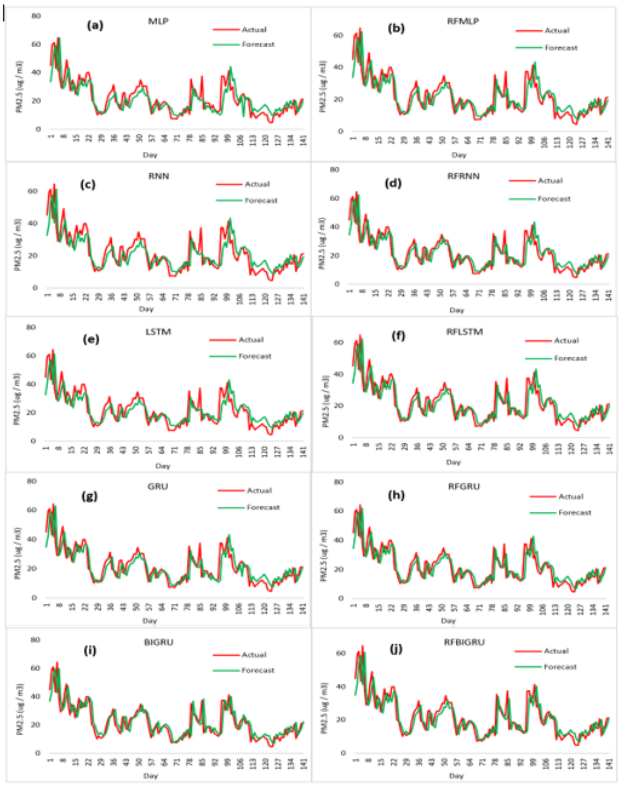


Figure 9. PM2.5 Forecasting Results of ADS without Feature Selection and with Feature Selection

Figure 10 proves the important role played by feature selection in improving PM2.5 forecasting results of MLP, RNN, LSTM, GRU and BIGRU. Among all, the PM2.5 forecasts of the proposed RFBIGRU (blue color) is much closer to actual PM2.5 (red color) compared to all other models. The selected features of the random forest helped to improve the bidirectional gated recurrent unit performance in forecasting PM2.5 for ADS. The performance of the proposed RFBIGRU model is also tested by using RMSE.

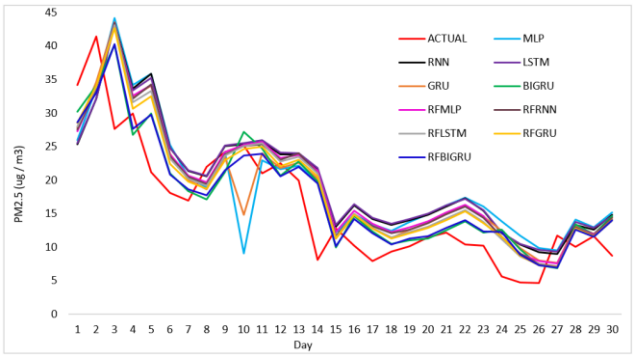


Figure 10. Sample Comparison of PM2.5 Forecasting Results of ADS

Table 5 shows the comparison of the forecasting results of DDS and ADS PM2.5 prediction models in terms of RMSE. The RFBIGRU outperformed other models by producing the least RMSE of 6.813. It also proves that all models with random forest selected features achieved better results than the models with all actual features. The MLP with selected features produced an RMSE of 6.880 which is less than 7.258 of the MLP with all actual features. The RNN with random forest selected features achieved an RMSE of 6.861 which is less than 7.089 of the RNN with all actual features. The LSTM with random forest selected features produced an RMSE of 6.843 which is less than 7.025 of LSTM with all actual features. The GRU with random forest selected features achieved a good result with the RMSE of 6.835 which is less than 6.936 of the GRU with all actual features. The BIGRU with random forest selected features produced remarkable PM2.5 forecasting results with an RMSE of 6.813 which is less than 6.901 of BIGRU with all actual features. In addition to that, the Friedman test is also performed and RFBIGRU performance is compared against all others.

Table 5. Forecasting Results of DDS and ADS in terms of RMSE

Methodology	RMSE	
	DDS	ADS
MLP	43.709	7.258
RNN	43.147	7.089
LSTM	43.071	7.025
GRU	42.948	6.936
BIGRU	42.907	6.901
RFMLP	42.800	6.880
RFRNN	42.707	6.861
RFLSTM	42.583	6.843
RFGRU	42.428	6.835
RFBIGRU	42.217	6.813

Table 6 and Table 7 provides Friedman test results for DDS and ADS datasets respectively. The generality of RFBIGRU is evinced using Friedman test for both DDS and ADS datasets. Table 6 shows the nonparametric Friedman test RESULTS for DDS. The RFBIGRU compared against all other models with the significance level as 0.10, 0.02 and 0.05.

Table 6. Friedman Test Results for DDS

Compared Models	Significance Level		
	α=0.10	α=0.02	α=0.05
RFBIGRU vs. MLP			
RFBIGRU vs. RNN			
RFBIGRU vs. LSTM	H_0: $e_1 = e_2 = e_3 = e_4 = e_5 = e_6 = e_7 = e_8 = e_9 = e_{10}$	H_0: $e_1 = e_2 = e_3 = e_4 = e_5 = e_6 = e_7 = e_8 = e_9 = e_{10}$	H_0: $e_1 = e_2 = e_3 = e_4 = e_5 = e_6 = e_7 = e_8 = e_9 = e_{10}$
RFBIGRU vs. GRU			
RFBIGRU vs. BIGRU			
RFBIGRU vs. RFMLP	F= 326.27 p=0.00000 (Reject H_0)	F= 326.27 p=0.00000 (Reject H_0)	F= 326.27 p=0.00000 (Reject H_0)
RFBIGRU vs. RFRNN			
RFBIGRU vs. RFLSTM			
RFBIGRU vs. RFGRU			

In all cases, the null hypothesis is rejected. It proves that the proposed model achieved a better PM2.5 forecasting result than all other models. Table 7 shows that the Friedman test is conducted between RFBIGRU and all other models for ADS dataset with the significance levels as 0.10, 0.02 and 0.05. All the three cases demonstrates that the null hypothesis is rejected. It evinces the superior performance of RFBIGRU against all others.

Table 7. Friedman Test Results for ADS

Compared Models	Significance Level		
	α=0.10	α=0.02	α=0.05
RFBIGRU vs. MLP			
RFBIGRU vs. RNN			
RFBIGRU vs. LSTM	H_0: $e_1 = e_2 = e_3 = e_4 = e_5 = e_6 = e_7 = e_8 = e_9 = e_{10}$	H_0: $e_1 = e_2 = e_3 = e_4 = e_5 = e_6 = e_7 = e_8 = e_9 = e_{10}$	H_0: $e_1 = e_2 = e_3 = e_4 = e_5 = e_6 = e_7 = e_8 = e_9 = e_{10}$
RFBIGRU vs. GRU			
RFBIGRU vs. BIGRU			
RFBIGRU vs. RFMLP	F= 16.44 p=0.00000 (Reject H_0)	F= 16.44 p=0.00000 (Reject H_0)	F= 16.44 p=0.00000 (Reject H_0)
RFBIGRU vs. RFRNN			
RFBIGRU vs. RFLSTM			
RFBIGRU vs. RFGRU			

5. Conclusion

In this paper, the highly hazardous PM2.5 pollutant in the air is predicted using the RFBIGRU model to save communities' lives by protecting them from severe health issues. A precise prediction of PM2.5 is achieved by RFBIGRU by reducing challenges like incompleteness, the curse of dimensionality, overfitting and improving the accuracy by using random forest feature importance as feature selector and bidirectional gated recurrent unit deep neural network for analyzing the temporal sequence patterns in the historical air pollutant data. The model's generality is also tested by using two different characteristic air pollutant data collected from two different regions namely Delhi and Amaravathi in India. The proposed RFBIGRU performance is also compared with MLP, RNN, LSTM, GRU, BIGRU, RFMLP, RFRNN, RFLSTM and RFGRU by measuring root mean square error. RFBIGRU produces the least RMSE 42.217 for the Delhi region and 6.813 for the Amaravathi region. The Friedman statistical test is also conducted and evinced the superiority of RFBIGRU. This research still has a future scope for enhancing the PM2.5 prediction accuracy by considering meteorological and population data and also incorporating the nature-inspired optimization techniques.

References

Alegria, A., Barbera, R., Boluda, R., Errecalde, F., Farré, R. and Lagarda, M.sJ. (1991). Environmental cadmium, lead and nickel contamination: possible relationship between soil and vegetable content, *Fresenius' Journal of Analytical Chemistry*, **339**, 654–657.

Aliyu, Y.A., Musa, I.J. and Jeb, D.N. (2014). Geostatistics of pollutant gases along high traffic points in urban Zaria, Nigeria, *International Journal of Geomatics and Geosciences*, ***5*(1)**, 19–31.

Amelia, R., Mahardika, R.G., Niani, C.R. and Lewaherilla, N. (2022). Predicting particulate matter PM2. 5 using the exponential smoothing and Seasonal ARIMA with R studio, In IOP Conference Series: Earth and Environmental Science, IOP Publishing, **1108(1)**, 012079.

Andre Gensler. (2019). Wind Power Ensemble Forecasting: Performance Measures and Ensemble Architectures for

Deterministic and Probabilistic Forecasts, kassel university press GmbH.

Athira, V., Geetha, P., Vinayakumar, R. and Soman, K.P. (2018). Deepairnet: Applying recurrent networks for air quality prediction, *Procedia computer science*, **132**, 1394–1403.

Ayturan, Y.A., Ayturan, Z.C. and Altun, H.O. (2018). Air pollution modelling with deep learning: a review, *International Journal of Environmental Pollution and Environmental Modelling*, **1(3)**, 58–62.

Bedekar, G., Patil, R.S., Tergundi, P. and Goudar, R.H. (2021). An Efficient Implementation of ARIMA Technique for Air Quality Prediction, Available at SSRN 3889537.

Bekkar, A., Hssina, B., Douzi, S. and Douzi, K. (2021). Air-pollution prediction in smart city, deep learning approach, *Journal of big Data*, **8(1)**, 1–21.

Bouktif, S., Fiaz, A., Ouni, A. and Serhani, M.A. (2020). Multi-sequence LSTM-RNN deep learning and metaheuristics for electric load forecasting, *Energies*, **13(2)**, 391.

Box, G.E., Jenkins, G.M., Reinsel, G.C. and Ljung, G.M. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.

Chen, R.C., Dewi, C., Huang, S.W. and Caraka, R.E. (2020). Selecting critical features for data classification based on machine learning methods. *Journal of Big Data*, **7(1)**, 52.

Ding, C., Zheng, Z., Zheng, S., Wang, X., Xie, X., Wen, D., Zhang, L. and Zhang, Y. (2022). Accurate air-quality prediction using genetic-optimized gated-recurrent-unit architecture, *Information*, **13(5)**, 223.

Dong, Y., Zhang, Z. and Hong, W.C. (2018). A hybrid seasonal mechanism with a chaotic cuckoo search algorithm with a support vector regression model for electric load forecasting, *Energies*, **11(4)**, 1009.

Drewil, G.I. and Al-Bahadili, R.J. (2022). Air pollution prediction using LSTM deep learning and metaheuristics algorithms, *Measurement: Sensors*, **24**, 100546.

Ercilla-Montserrat, M., Muñoz, P., Montero, J.I., Gabarrell, X. and Rieradevall, J. (2018). A study on air quality and heavy metals content of urban food produced in a Mediterranean city (Barcelona), *Journal of cleaner production*, **195**, 385–395.

Ghiasi, M., Irani Jam, M., Teimourian, M., Zarrabi, H. and Yousefi, N. (2019). A new prediction model of electricity load based on hybrid forecast engine, *International Journal of Ambient Energy*, **40(2)**, 179–186.

Ghose, B., Rehena, Z. and Anthopoulos, L. (2022). A Deep Learning based Air Quality Prediction Technique Using Influencing Pollutants of Neighboring Locations in Smart City, *JUCS: Journal of Universal Computer Science*, **28(8)**.

Guo, Z., Yang, C., Wang, D. and Liu, H. (2023). A novel deep learning model integrating CNN and GRU to predict particulate matter concentrations, *Process Safety and Environmental Protection*, **173**, 604–613.

Hong, W.C. ed. (2018). *Hybrid Advanced Optimization Methods with Evolutionary Computation Techniques in Energy Forecasting*. MDPI.

Huang, G., Li, X., Zhang, B. and Ren, J. (2021). PM2. 5 concentration forecasting at surface monitoring sites using GRU neural network based on empirical mode decomposition, *Science of the Total Environment*, **768**, 144516.

Kim, B., Kim, E., Jung, S., Kim, M., Kim, J. and Kim, S. (2023). PM2. 5 Concentration Forecasting Using Weighted Bi-LSTM and Random Forest Feature Importance-Based Feature Selection, *Atmosphere*, **14(6)**, 968.

Kiruthika, V.G., Arutchudar, V. and Senthil Kumar, P. (2014). Highest humidity prediction using data mining techniques, *Int J Appl Eng Res*, ***9*(16)**, 3259–3264.

Kumar, K. and Pande, B.P. (2022). Air pollution prediction with machine learning: a case study of Indian cities, *International Journal of Environmental Science and Technology*, 1–16.

Kumar, U. and Jain, V.K. (2010). ARIMA forecasting of ambient air pollutants (O 3, NO, NO 2 and CO), *Stochastic Environmental Research and Risk Assessment*, **24**, 751–760.

Li, C., Hsu, N.C. and Tsay, S.C. (2011, A study on the potential applications of satellite data in air quality monitoring and forecasting, *Atmospheric Environment*, **45(22)**, 3663–3675.

Li, Q., Zhu, Q., Xu, M., Zhao, Y., Narayan, K.V. and Liu, Y. (2021). Estimating the impact of COVID-19 on the PM2. 5 levels in China with a satellite-driven machine learning model, *Remote sensing*, **13(7)**, 1351.

Li, X., Peng, L., Yao, X., Cui, S., Hu, Y., You, C. and Chi, T. (2017). Long short-term memory neural network for air pollutant concentration predictions: Method development and evaluation, *Environmental pollution*, **231**, 997–1004.

Lin, K.P., Pai, P.F. and Yang, S.L. (2011). Forecasting concentrations of air pollutants by logarithm support vector regression with immune algorithms, *Applied Mathematics and Computation*, **217(12)**, 5318–5327.

Liu, X., Wang, Y., Wang, X., Xu, H., Li, C. and Xin, X. (2021). Bi-directional gated recurrent unit neural network based nonlinear equalizer for coherent optical communication system, Optics Express, **29(4)**, 5923–5933.

Lu, G., Yu, E., Wang, Y., Li, H., Cheng, D., Huang, L., Liu, Z., Manomaiphiboon, K. and Li, L. (2021). A novel hybrid machine learning method (OR-ELM-AR) used in forecast of PM2. 5 concentrations and its forecast performance evaluation. *Atmosphere*, **12(1)**, 78.

Ma, J., Ding, Y., Gan, V.J., Lin, C. and Wan, Z. (2019). Spatiotemporal prediction of PM2. 5 concentrations at different time granularities using IDW-BLSTM, *Ieee Access*, **7**, 107897–107907.

Manan, W.D.W.A. and Abdullah, N.R. (2021). An artificial neural network model for forecasting air pollution, In *IOP Conference Series: Materials Science and Engineering,* IOP Publishing, **1176(1)**, 012032.

Mani, G. and Viswanadhapalli, J.K. (2022). Prediction and forecasting of air quality index in Chennai using regression and ARIMA time series models, *Journal of Engineering Research*, **10(2A)**, 179–194.

Masood, A. and Ahmad, K. (2020). A model for particulate matter (PM2. 5) prediction for Delhi based on machine learning approaches, *Procedia Computer Science*, **167**, 2101–2110.

Obando Bobadilla, L.M., Ruiz Nieto, M.P. and Rodríguez Molano, J.I. (2018). From traditional cities to smart cities, In *Knowledge Management in Organizations: 13th International Conference, KMO 2018, Žilina, Slovakia,* Springer International Publishing, *Proceedings* **13***,* 661–673.

Pruthi, D. and Liu, Y. (2022). Low-cost nature-inspired deep learning system for PM2. 5 forecast over Dsselhi, India. *Environment International*, **166**, 107373.

Qing, L. (2023). PM2. 5 Concentration Prediction Using GRA-GRU Network in Air Monitoring, *Sustainability*, ***15*(3)**, 1973.

Saiohai, J., Bualert, S., Thongyen, T., Duangmal, K., Choomanee, P. and Szymanski, W.W. (2023). Statistical PM2. 5 Prediction in an Urban Area Using Vertical Meteorological Factors. *Atmosphere*, **14(3)**, 589.

Senthil Kumar, P. (2017). A review of soft computing techniques in short-term load forecasting, *Int J Appl Eng Res*, ***12*(18)**, 7202–7206.

Siva Sankari, S. and Senthil Kumar, P. (2023). A Review of Deep Transfer Learning Strategy for Energy Forecasting, *Nature Environment and Pollution Technology*, **22(4)**, 1781–1793.

Subbiah, S.S. and Chinnappan, J. (2020). A review of short term load forecasting using deep learning, *International Journal on Emerging Technologies*, **11(2)**, 378–384.

Subbiah, S.S. and Chinnappan, J. (2021). Opportunities and Challenges of Feature Selection Methods for High Dimensional Data: A Review, *Ingénierie des Systèmes d'Information*, **26(1)**, 67–77.

Subbiah, S.S. and Senthil Kumar, P. (2022). Deep learning based load forecasting with decomposition and feature selection techniques, *Journal of Scientific & Industrial Research*, **81(5)**, 505–517.

Subbiah, S.S., Paramasivan, S.K., Arockiasamy, K., Senthivel, S. and Thangavel, M. (2023). Deep Learning for Wind Speed Forecasting Using Bi-LSTM with Selected Features. *Intelligent Automation & Soft Computing*, **35(3)**, 3829–3844.

Swaroop, G., Senthil Kumar, P. and Muthamil Selvan, T. (2014). An efficient model for share market prediction using data mining techniques, *Int J Appl Eng Res*, ***9*(17)**, 3807–3812.

Vijayanti, V., Kar, A. and Navya, K.N.S.S. (2023). Analysis of Deep Learning Based Garbage Detection in Water Bodies, In *2023 5th International Conference on Inventive Research in Computing Applications (ICIRCA),* IEEE, 1–8.

Wang, P., Liu, Y., Qin, Z. and Zhang, G. (2015). A novel hybrid forecasting model for PM10 and SO2 daily concentrations. *Science of the Total Environment*, **505**, 1202–1212.

Xiao, Q., Chang, H.H., Geng, G. and Liu, Y. (2018). An ensemble machine-learning model to predict historical PM2. 5 concentrations in China from satellite data, *Environmental science & technology*, **52(22)**, 13260–13269.

Xiaogang, L. and Xin, L. (2022). Air Quality Index Prediction Based on Multilayer Perceptron, *Academic Journal of Computing & Information Science*, **5(10)**, 101–105.

Xiaowei, W. (2022). Air Quality Level Prediction Based on Multilayer Perceptron and BP Back Propagation Algorithm, In *2022 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA),* IEEE, 620–624.

Xu, H., Zhang, A., Xu, X., Li, P. and Ji, Y. (2022). Prediction of Particulate Concentration Based on Correlation Analysis and a Bi-GRU Model, *International Journal of Environmental Research and Public Health*, **19(20)**, 13266.

Yu, R., Yang, Y., Yang, L., Han, G. and Move, O.A. (2016). RAQ–A random forest approach for predicting air quality in urban sensing systems, *Sensors*, **16(1)**, 86.

Zhan, Y., Luo, Y., Deng, X., Chen, H., Grieneisen, M.L., Shen, X., Zhu, L. and Zhang, M. (2017). Spatiotemporal prediction of continuous daily PM2. 5 concentrations across China using a spatially explicit machine learning algorithm. *Atmospheric environment*, **155**, 129–139.

Zhao, X., Zhang, R., Wu, J.L. and Chang, P.C. (2018). A Deep Recurrent Neural Network for Air Quality Classification, *J. Inf. Hiding Multim. Signal Process,* **9(2)**, 346-–354.