ECTI Transactions on Computer and Information Technology
Journal homepage: https://ph01.tci-thaijo.org/index.php/ecticit/
Published by the ECTI Association, Thailand, ISSN: 2286-9131

Improving Air Quality Prediction with a Hybrid Bi-LSTM and GAN Model

Rupa Rajakumari R Peter[1] and Ujwal Ambadas Lanjewar[2]

ABSTRACT

Air quality is a topic that has been of utmost concern across the globe for the past few decades. Various intelligent monitoring systems are used in diverse scenarios, collecting air quality data that contains missing values. Such missing values in data cause hindrances in forecasting. This time series prediction or forecasting process extracts the necessary information from historical records and predicts future values. To solve the missing values issue in data, Generative Adversarial Networks (GAN) are used to impute the missed data. While the learning of long-term dependencies embedded in the time series poses another threat to the models in the time prediction. To overcome this, Long Short-Term Memory (LSTM) models are used. Yet, most of the neural network-based methods failed to consider the patterns of time series data that varied for each period, and the encoder-decoder performance deteriorated for longer sequences. To combat this, the present study proposes a hybrid probabilistic model to generate parameters for predictive distribution at every step. Hence, an implementation of hierarchical-attention-based BiLSTM with GAN is proposed in the study for effective prediction and minimal error. The proposed model is assessed with the evaluation metrics such as Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), and Mean Square Error (MSE). The evaluation metric confirmed the higher accuracy of the proposed model than the existing models in time series prediction.

Article information:

Keywords: Air Quality Data, Time Series Prediction, Hierarchical-attention Based BiLSTM, GAN, Accuracy

Article history:
Received: February 15, 2023
Revised: April 10, 2023
Accepted: August 11, 2023
Published: September 2, 2023 (Online)

DOI: 10.37936/ecti-cit.2023173.251670

1. INTRODUCTION

Air pollution is regarded as the air getting contaminated due to certain harmful substances that result in serious health issues for humans and other living beings on Earth. It also causes serious damage to the materials and climate. Pollutants that pollute air include gases (like carbon monoxide, nitrous oxides, ammonia, and carbon dioxide), biological molecules, and particulates (both inorganic and organic). Air pollution can cause allergies, diseases, and even death to human beings. Air pollution has significant factors associated with pollution-related diseases such as heart diseases, respiratory diseases, lung cancer, and stroke. Poor air quality affects human health, affecting the body's respiratory and cardiovascular systems. NCAP reports stated that around 43 cities in India come under 102 nonattainment cities of India. More than half of the country's population is exposed to poor-quality air, which approximately exceeds the considerable limits. Recent research on global air pollution has reported that almost 600000 premature deaths are encountered annually in India due to air pollution [1]. Air pollution control in India has become tough due to the great impact of meteorological factors and traffic emissions, which is even more difficult to investigate. Considering the serious impacts of air pollution, the government has taken many smart leads. It is currently working with many research in-

[1] The author is with Research Scholar, Department of Electronics and computer Science, Rashtrasant Tukadoji Maharaj Nagpur University, Nagpur, India, E-mail: ruparajakumari@gmail.com
[1] The author is with Department of Computer Science, Hislop College, Nagpur, India, E-mail: ruparajakumari@gmail.com
[2] The author is with Principal, Smt. Binzani Mahila Mahavidyalaya, Nagpur, India, E-mail: ualanjewar@gmail.com
[1] The corresponding author: ruparajakumari@gmail.com

stitutions to take the necessary actions to control the ambient levels of air pollution. Many air pollution monitoring stations were developed by the government to collect the quality of air that can be used for forecasting the next hours, day, or week. The forecasting outcomes provide information if there's any need for preventive activities in advance. Thus, air quality monitoring and modeling can highly assist in alleviating the severe impact of air pollution. Many technologies have been employed to develop the air prediction models, such as statistical, deterministic, deep learning, and machine learning.

To combat such problems, several investigators initially used machine learning frameworks like random forest [2], support vector machine [3], fuzzy neural network[4], artificial neural networks [5], linear regression, and XGBoost [6]. Among these models, feed-forward neural network (NN) based Artificial Neural Network (ANN) has revealed improved performance in prediction. Despite showing better accuracy in air quality prediction, these neural network models do not investigate the correlation among the features of the multivariate air pollution dataset. To rectify this limitation, time series forecasting models were considered and employed. These models help foresee the future value of the given target t for a given entity i at a time of t. All the entities in the model represent the logical grouping data, such as dimensions fetched from an individual weather station in the climatology or some important signs from various patients in medicinal fields. With rapid technological improvement, machine learning models were replaced by deep learning algorithms. Researchers started to utilize deep learning models for air quality modeling and have been proven to possess a better quality in prediction when compared with machine learning models about the temporal analysis of the air pollution dataset. In addition, deep learning algorithms performed better in human detection, medical image classification, sequential modeling, and many more. Particularly, deep learning models such as LSTM (Long Short Term Memory), RNN (Recurrent Neural Network), and Gated Recurrent Unit (GRU) frameworks have a significant role to play in forecasting.

However, for an RNN framework, the impact of any given input on the hidden layers and eventually on the neural network output would either blow up or decay exponentially when cycling around recurrent connections. Also, without customized hardware and software accelerations, LSTM computing time is directly proportional to the number of parameters. This is one disappointing feature of LSTMs. Thus, to overcome certain limitations of LSTMs, hierarchical-attention-based BiLSTMs are used. This is because hierarchical-attention-based BiLSTMs are capable of capturing the underlying context in a better way as it traverses the input data twice [7]. Also, LTSM-based models have the issue of remembering the problem of long data sequences. This is because the LSTM models are traversed only once from left to right, some inputs will be fed training model. Whereas the hierarchical-attention-based BiLSTM-based model trains the input data forwardly (from left to right) and backwardly (from right to left). Thus, the length of the training data is managed through each batch which is half of the data learned through each batch by LSTM. While predictive models show considerable accuracy in prediction, the estimation of missing values remained a challenge in the prediction. Hence, imputation algorithms have been used to find the missing values in the data that was measured or observed. The imputation method, the generative adversarial networks (GAN), accurately determines the missing data and gives a generalized output. Thus, the present study has considered the advantage of hierarchical-attention-based BiLSTM and GAN in time series prediction for weather forecasting. The main contributions of the work are as follows,

- To develop a time series data prediction model for one-step-ahead and multi-horizon time series forecasting using hierarchical-attention-based BiLSTM and GAN model to improve the predictive accuracy and minimize error rates.
- To overcome the data imbalance problem of time series data with the data imputation algorithm GAN.
- To achieve better prediction accuracy and high performance in terms of Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), and Mean Square Error (MSE) metrics.
- To compare the proposed model with the other existing benchmark models in time series prediction.

1.1 Significant contribution

The proposed model combines a bi-directional LSTM and GAN model for handling the missing data. The model's training is done by utilizing a hierarchical attention-based technique in which the input data is passed into an attention layer that stresses the significance of every time step in the input sequence. During the training process, the bi-directional LSTM is initially trained to forecast the subsequent value in time series based on available input data. The GAN model is then utilized to impute any missing values in the input sequence to apprise the LSTM model weights. This procedure is reiterated till convergence is attained. The Proposed model is initially trained by utilizing the training dataset, which involves incomplete and complete time series data. The GAN model is utilized to impute any missing values in training data that is utilized for training the LSTM model. After the training process by the model is completed, it is utilized to create predictions on unidentified time series data. Throughout the prediction, the model considers the input as the order of former values and utilizes the LSTM model to forecast subsequent values in time series. If there are missing values in the input sequence, before making the forecasting, the GAN model is utilized to impute it. This procedure is reiterated to produce the sequence of the predicted values. The proposed hybrid model encounters this limitation by utilizing the GAN model to impute missing values. This, in turn, enhances the accuracy of the predictions. The proposed model uses a bi-directional LSTM model for time series prediction and is integrated with the GAN model for handling missing data. A hierarchical attention-based approach is utilized to train the model. Hence, the proposed model can be used to predict unidentified time-series data.

2. LITERATURE REVIEW

The following section of the literature review discusses the various time series forecasting algorithms utilized in the existing methods for future predictions.

Air pollution is considered a major problem in the living environment since air quality is affected by various factors, like the smoke from industries due to power generators and vehicle traffic. The polluted air is mixed up with numerous harmful particles that affect physical health badly and threatens the life of humans. Therefore, forecasting air quality is important and has been regarded as the primary issue in protecting the environment. Several monitoring systems with deep learning (DL) have been designed for forecasting air quality to control air pollution. The suggested study[8] has designed a model for forecasting air quality. It has been formulated to learn spatial and temporal correlation features and multivariate time series data by combining two DL models, Convolutional Neural Network (CNN) and Bi-LSTM (Bidirectional Long Short Term Memory) networks. The time series data are generally dynamic and non-linear. The local and spatially correlated features have been extracted through CNN followed by Bi-LSTM operation, which has learned the dependencies between the spatial and temporal data. The system has adequate potential to predict air quality, but there is a lack of determining changes in the time series data.

Certain air pollution data outliers must be managed in advance through various forecasting conditions. So, the study [9] has concentrated on the pollution control forecasting operation before future situations, which efficiently helps prevent air pollution. The concentration of air pollutants dynamically changes based on wind speed, temperature, rainfall, the direction of the wind, and snowfall which makes it difficult to predict. The recommended study used PM2.5 (Particulate Matter with 2.5 μm in diameter) and CBGRU (CNN-based Gated Recurrent Unit) for future predictions. The feature learning abilities have been improved with the considered methods, and hence the system has achieved better prediction through the utilization of content being integrated with the CBGRU. The deep learning-based Bi-LSTM model has provided the ability to learn from the long-term dependencies, and the study [10] has applied transfer learning for transferring the knowledge learned from the smaller resolutions to greater temporal resolutions. The concentration of pollutants in the air has been predicted at various time resolutions. The model has been performed by monitoring the hourly concentration of data and has analyzed air quality with certain conditions. Air quality monitoring with different geographical conditions could prevent pollution better.

The poor quality of air can lead to numerous negative impacts on health, which have motivated the recent developments of systems in deep learning with the time series prediction of data, and the recommended study[11] has utilized a Bayesian-based DL framework for forecasting air pollution. For capturing the historical data on weather, the considered method has included an attention layer mechanism which also has been capable of determining the recursive spatial and temporal correlation of collected air quality data. The model has reduced prediction errors, and the fusion of the Bayesian method into the forecasting operation and feature selection methods has improved the interpretability factor with better prediction. Analyzing time series data for predicting the quality and fineness of air is considered a challenging issue since those kinds of data are generally partially missing, particularly when it has been collected simultaneously from various locations. The missing value-based imputation method has been designed in the suggested research [12] for solving the missing values. Since the missing values were difficult to interpret, analysis with the imputed data has also been difficult.

The DL-based technique has helped impute missing values for both spatial and temporal data. During data collection and analysis, data missing occurs due to outliers or anomalies. If such missing values because of anomalies have not been treated, it leads to issues in regression or classification-related tasks. Numerous methods for time series prediction have lacked dealing with the anomalies in the data being collected, which affects the prediction accuracy eventually. Imputation techniques have either imposed stronger assumptions on the missing value distribution or have not been applied completely. The feature correlations and the temporal dependencies at various time frequencies have been ignored. The RNN (Recurrent Neural Network) model has been designed in the study [13] to generate the missing values from the observed data with both temporal and non-temporal information. The discriminator has been designed to measure values being generated, and it has the responsibility of adjusting the generator to provide imputed output. However, under the higher level of the missing rate, the pattern of missing time series data has to provide information to the generator for updating the system. The effective handling of missing values and attaining prediction accuracy has been lacking in existing methods. Hence, the suggested methodology in [14] has utilized a GAN model to predict time series data using deep learning. The recommended model has predicted weather parameters such as precipitation, temperature, and pressure to predict whether the air is polluted. The study has analyzed time series data for only specific meteorological weather parameters, and uncertainty in weather conditions may reduce accurate prediction. The GAN-based learning method deals with the time series data for the complete exploitation of interdependencies within the time steps and correlations among variables in a single step. It has the potential to capture the patterns of time series and learn from the long-term dependencies of air pollutant data.

It is important to formulate a mechanism or scheme through which the air pollutants can be predicted since the air quality affects people's lives to a greater extent. Similarly, another study[15] introduced the DL framework for forecasting air quality. The experimental outcomes of the investigation infer that the Bi-LSTM and LSTM algorithms, among other base algorithms, can provide highly precise long-term predictions of air quality. About this, a study[16] performs the forecasting of air pollutants using ML and DL models such as LSTN encoder-decoder, LSTM, and FbProphet. The study predicted the hourly and daily Particulate matter 2.5 ($PM_{2.5}$) Concentration. The study outcomes reveal that the LSTM encoder-decoder outperforms other methods in forecasting air pollutants ($PM_{2.5}$).

Problem Identification

The existing studies lacked certain limitations in which the common issues are identified and listed as follows:

- The prediction model for forecasting air pollution has lacked the detection and management of outliers in time series data and can improve the model to operate under uncertain conditions in forecasting [8].
- The Bayesian method for forecasting has provided accuracy in forecasting only on integration with various forecasting methods and lacked the inclusion of factors like spatial and temporal data from nearer monitoring stations which could lower the air quality prediction [11].

3. PROPOSED METHODOLOGY

Air quality data are multivariate time series and are continuing that contain time measurements where the current measurement will be somewhere related to the previous measurement of data and hence dependent. Time series data are generally impacted by the non-stationarity and collinearity, leading to the violation of independent assumptions. This ultimately leads to difficulty in forecasting. To overcome this, the proposed study uses a hybrid probabilistic model that uses NN to generate the parameters for prognostic distributions at every step. The proposed model is a hybrid model that combines two deep learning algorithms, namely the hierarchical-attention-based bidirectional LSTM (Bi-LSTM) with Generative Adversarial Networks (GAN) for time series forecasting based on temporal data. In that, attention mechanisms assist in selecting required information among different feature time series data to predict the target time series. In addition to these, the proposed model also uses the Adam optimization algorithm as the time-varying optimizer. This hybrid model combines the advantages of both algorithms, where the classification memory ability and superior memory of LSTM and the generative and discriminative qualities of GAN complement the model with better time series prediction. While the computation time is reduced by the Adam optimization algorithm as it optimizes the parameters of the hierarchical-attention-based Bi-LSTM. The proposed hybrid model is collectively known as the BiLSTM-GAN model, where the layers of BiLSTM and GANs are used to construct the proposed encoder-decoder model. The overall view of the proposed model is presented below in Figure 1.

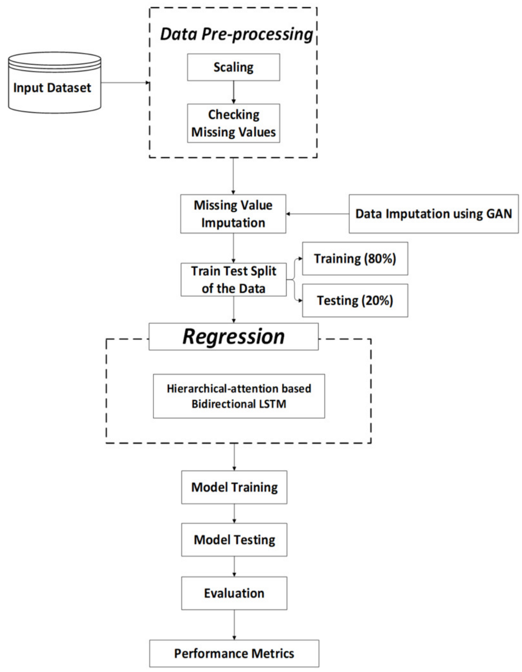

Fig.1: *Overall View of the Proposed System.*

Initially, the air pollution dataset is loaded. Then the data is pre-processed for scaling and checking the missing values. Followed by that, the proposed data imputation using GAN is done to impute the missing values. Followed by it, the train test split of the data takes place. The present study employs the most commonly used ratio of 80:20 for data split into training and testing. The proportion of training and testing data utilized in the study relies on various factors, like the complexity of the problem, dataset size, and the available computational resources. Then, Bi-LSTM receives the imputed values, and finally, prediction is done by the trained model. The performance metrics are used for the performance analysis of the proposed system.

3.1 Pre-Processing

Time series data contain numerous information which is not visible generally. The most common issues with the time series dataset are missing values, unordered time stamps, noise, and outliers. Hence, data pre-processing is done to check for missing values, noise, and other inconsistencies in data before the algorithm execution. In this process, scaling is done initially for data normalization. Among the mentioned issues, the most difficult one is the missing value which is overcome in this study by the following algorithms discussed below.

3.2 Data Imputation using GAN

The module of data imputation in the proposed study relies on GAN, which learns the correlation between the input layer data to replace the missing measurements. GAN is usually optimized by assessing its performance grounded on the imputed measurements on detection modules and fault identification. Ultimately, the monitoring systems perform two vital tasks. In addition, the imputed data through GAN alleviates loss on the modules even in continuously missing measurements required for the monitoring system. This GAN can operate under conditions where the data is completely unavailable successfully. The presence of a generator in the GAN architecture helps impute the data accurately, while the discriminator distinguishes the imputed and observed components. Also, the classification loss is minimized with the changed discriminator, and the discriminator's misclassification rate is maximized with the trained generator. Hence, the adversarial process is used for such training of the networks. To ensure the result desirability of the adversarial process, the discriminator is provided with additional information known as hints. This ensures that samples are generated concerning the true data distribution. The architecture of GAN is presented in below figure 2.

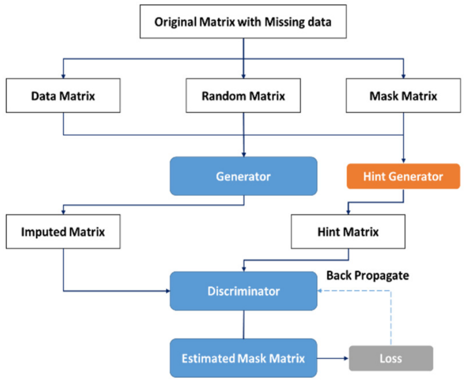

Fig.2: *GAN Architecture for Data Imputation.*

The prime purpose of data imputation in the proposed system is to replace the missing values so the classification modules operate accurately. To accomplish this, the model involves both faulty and non-faulty data in the process of training. For this purpose, the functions of the generator and discriminator are mandatory. The process of both is briefed below.

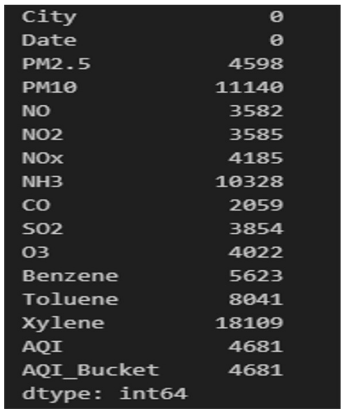

Fig.3: *Missing value Imputation using GAN.*

Table 1: *GAN Imputation vs. Mean Imputation in the proposed model.*

Error Metric	Mean Imputation	With GAN imputation
MSE	2247.26	44.24965212
MAE	2147.45	20.6984
MAPE	54971.2	0.1126676
RMSE	4758.52	6.652041

Figure 3 depicts the filling of missing values by GAN imputation. Table 1 shows that the error values of the proposed model with GAN imputation are lower than the Mean Imputation method. This, in turn, proves the efficiency of GAN in managing and handling the missing data. Hence, the present study utilizes the GAN imputation in the proposed model.

3.2.1 Generator

The generator G notices the real data vector x of n-dimensions that contain the missing components. The mask that denotes the missing values of the input data is denoted by mk. The data of missing values are found by multiplying the mask MK with the whole input data. The missing components are imputed by the generator G and provide an imputed vector $\grave{x}$ as output.

Table 2: *Generator and Discriminator Networks Dimension.*

	Size of Input Layer	Size of Output Layer	Size of Hidden Layer
Generator-G	13	13	[26,26]
Discriminator-D	13	13	[13,13]

Table 2 denotes the dimensions of both the generator and discriminator networks, the G - generator and D - discriminator have one input layer, one output layer, and two hidden layers. The size of the input layer and output layer for G and D is 13, and the size of the hidden layer for G and D is [26, 26] and [13, 13]. The input of G contains present measurements and the missing values, the mask which denotes the missing dimensions, and the measurements for the last time steps L_G= 12. Therefore, considering this, the present study contains n= 13 sensors. Then, the discriminator input contains the present sensor dimension and the imputed values, i.e., the G output and the hinting vector.

Table 3: *GAN Hyperparameters.*

	Alpha	Hint Rate	Batch Size	♯ Epochs
Generator-G	100	0.9	128	500
Discriminator-D	100	0.9	128	500

Table 3 shows the chosen alpha parameter, hint rate, batch size, and epochs. The chosen alpha range for G is 100, and the Hint rate is 0.9, with the batch size 128 at 500 epochs. Likewise, the chosen alpha of D is 100, and the hint rate is 0.9 with 128 batch size at 500 epochs.

3.2.2 Discriminator

The imputed vector is received by the discriminator D. The vectors generated by G are completely received by the discriminator that tries to determine the components that were credited and observed. To be precise, discriminator D tries to predict the mask mk. D also receives a hint vector apart from the output of G. This hint discloses the fractional information regarding the missing values. Specifically, the hint vector discloses all the components of mask mk to D, except those chosen independently and randomly for every sample. Followed by this, the training of the GAN model is based on two steps where discriminator D is optimized with a fixed value of generator by utilizing the mini-batches of specific size MB_D. Now the discriminator D is trained by eq. 1, and L_d is defined in eq.2 as follows,

$$\frac{\min}{D} - \sum_{j=1}^{k_D} L_d\big(mk(j), \grave{m}k(j), p(j)\big) \tag{1}$$

$$\begin{aligned} L_d(mk, \grave{m}k, p) = \sum_{i:p_i=0} [&mk_i \log(\grave{m}k_i) \\ &+ (1 - mk_i)\log(1 - \grave{m}k_i)] \end{aligned} \tag{2}$$

The original mask associated with the mini batch's j^{th} sample is denoted as $mk(j)$, and the respective predicted mask is represented as $m\grave{k}(j)$. This means that the discriminator's output and $p(j)$ are the n-dimensional vectors where all the elements are equal to the value, except one element whose value will be 0 and whose position in the respective p(j) denotes the mask element's position, which is denoted as $mk(j)$ and is not provided to D as input. To be more precise, discriminator D is trained with eq. 2 only for the mask vector element, which is unfamiliar to the discriminator for the randomly chosen samples.

$$\frac{\min}{G} \sum_{j=1}^{k_G} L_G\left(mk(j), \grave{m}k(j), p(j)\right) + \alpha\ L_M(x(j), \grave{x}(j)) \tag{3}$$

Once the training for D is started, generator D is optimized according to eq. 3 concerning the mini-batch size, namely MB_G. Now G's cost function is the hyperparameter α with a weighted sum of two components. One corresponds to observed measurements L_M, and the other corresponds to the missing LG measurements shown in eq. 4 and eq. 5 below.

$$L_G(mk, \grave{m}k, p) = -\sum_{i:p_i=0} (1 - mk_i)\log(\grave{m}k_i) \tag{4}$$

$$L_{mk}(x, \grave{x}) = \sum_{i=1}^{d} mk_i(\grave{x}_i - x_i)^2 \tag{5}$$

Thus the optimization of discriminator D is done with the fixed generator G imputed values.

3.3 Hierarchical-attention based BiLSTM

The algorithm used in the proposed study uses hierarchical-attention-based BiLSTM, through which it achieves effective prediction of time series. An attention mechanism has been utilized to extract relevant input features at every time step, which refers to the hidden state in the previous encoder. The attention mechanism assists in the effective prediction and interpretation. The study uses the regression method, as presented in Figure.1, which predicts the output based on the input features fed into the model. The characteristics of sharing weights and local perception help minimize the number of parameters through which the learning model can perform efficiently. The imputed data from GAN is fed to the regression model to build the hierarchical-attention-based BiLSTM. The extended LSTM version is the hierarchical-attention-based BiLSTM that contains two units of LSTM that process the input data in both backward and forward directions. In the initial round of the process, LSTM is used in a forward direction upon the input, and in the next round, the reversal of the input is fed into the LSTM. Thus applying the concept of LSTM twice aids in enhancing the learning of the long-term dependencies that enhances the model's accuracy. The architecture of BiLSTM is presented in Figure 4 below.

Algorithm 1: Hierarchical-attention-based BiLSTM

1.Input the features
2.Imputed Data
  = GAN(miss_data_x,gan_parameters)
3.Build a Hierarchical Bi-LSTM Model
4.for e = 1 to num_iter do
5. Attention Mechanism with Hierarchical Method
6. Input Dim,time_steps
  a. Per = Permute (Inputs) – Dense Layer
  b. Single Attention Vector – Repeat Vector
  c. Hier_attn - Multiply()([inputs, Per])
  d. Train the model for a batch of the dataset from imputed data with Hier_attn
  e. Minimize the RMSE loss
7. end for
8. Test Stage
9. for e = 1 to num$_i$*terdo*
  a.Fetch the sample from the test
  b.Evaluate the predictive performance
10. end for

The representations $\overrightarrow{h_t}$ and $h_t^{\leftarrow}$ denote the output of the forward and backward input layer, respectively, and the input and output vectors are x_t and y_t, respectively. Now $\overrightarrow{h_t}$ is measured iteratively with the time range $t-n$ to $t-1$, and $h_t^{\leftarrow}$ is measured with reversed inputs of the same time range. The outputs of the backward and forward layers are calculated using the updating equations in eq. 7 to eq. 12. The output vector y_t is obtained by combining the two output sequences. Thus it is given in eq. 6 as,

$$y_t = \sigma(\overrightarrow{h_t}, h_t^{\leftarrow}) \tag{6}$$

Where the two output sequences are combined by σ.

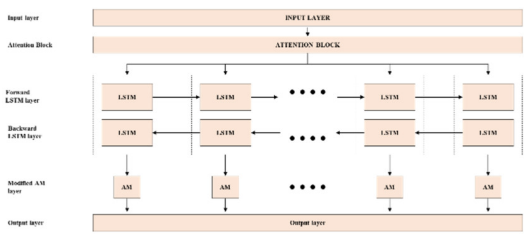

Fig.4: *Functional Framework of BiLSTM.*

The hierarchical-attention-based BiLSTM contains the forget gate denoted as f_t, which controls the removal of unnecessary information present in the memory cel $\hat{c}_t$. The input gate controls the addition that occurs due to new information. The output gate o_t is used to control the internal memory state exposure. The three gates help the memory cell to delete, forget and update internal information. In a given time t, the output gate o_t, forget gate f_t, input gate i_t memory cell $\hat{c}_t$, and hidden cell h_t transition functions are given below in the following equations.

$$f_t = \sigma_g(Wgt_f x_t + P_f h_{t-1} + b_t) \tag{7}$$

$$i_t = \sigma_g(Wgt_i x + P_i h_{t-1} + b_i) \tag{8}$$

$$o_t = \sigma_g(Wgt_o x + P_o h_{t-1} + b_o) \tag{9}$$

$$\hat{c}_t = \tan\ \tan\ h(Wgt_c x + P_c h_{t-1} + b_c) \tag{10}$$

Where Wgt_f, Wgt_i, Wgt_o, and Wgt_c denote the weight matrices that map the input of the hidden layer to the input cell state and three gates. While P_f, P_i, P_o, and P_c denote the weight matrices that connect the output state of the previous cell to the input cell and the three gates. The terms b_t, b_i, b_o, and b_c denote the bias factors. The gate activation is denoted by σ_g, the sigmoid function and the hyperbolic tangent function is denoted by as $\tan\ \tan\ h$. Thus based on the previous equations, the results at hidden layer output, cell output state at time t are calculated as,

$$c_t = f_t * c_{t-1} + i_t * c_t \tag{11}$$

$$h_t = o_t * \tan\ \tan\ h(c_t) \tag{12}$$

The neural network with the drop-out and arbitrary depth is provided at each weight layer and is mathematically equal to the probabilistic model. This is achieved by model training with the dropout after every hidden layer. In the proposed work, dropout is provided with the BiLSTM layer to minimize data overfitting. The additional layers of the BiLSTM aid in training the input data twice and also help to learn the uncertainties and fine granularity to improve the predictive accuracy table 4 shows the trainable parameters. While the dropout regularization aids in minimizing the model capacity to attain a lower generalization error. The neurons dropped randomly at various points during training due to the dropout probability at each epoch. This prevents overfitting and combines various neural networks.

Table 4: BiLSTM Layer and Parameters.

Layer	Output shape	Parameter
input_1 (Input Layer)	(None, 12, 1)	0
conv1d (Conv1) ['input_1[0][0]']	(None, 12, 64)	128
dropout (Dropout) ['conv1d[0][0]']	(None, 12, 64)	0
bidirectional (Bidirectional) ['dropout[0][0]']	(None, 12, 128)	66048
dropout_1 (Dropout) ['bidirectional[0][0]']	(None, 12, 128)	0
dense (Dense) ['dropout_1[0][0]']	(None, 12, 128)	16512
attention_vec (Permute) ['dense[0][0]']	(None, 12, 128)	0
multiply (Multiply) ['dropout_1[0][0]', 'attention_vec[0][0]']	(None, 12, 128)	0
flatten (Flatten) ['multiply[0][0]']	(None, 1536)	0
dense_1 (Dense) ['flatten[0][0]']	(None, 1)	1537
Total Parameters		84,225
Trainable Parameters		84,225
Non-Trainable Parameters		0

4. RESULTS AND DISCUSSION

The proposed model's efficiency and effectiveness are analyzed regarding RMSE, MAE, and MAPE using Air Quality Data in India (2015 - 2020). The obtained results are enumerated in the following subsections.

4.1 Dataset Description

The experimental evaluation of the proposed model is performed by using Air Quality Data in India (2015-2020), which is openly available on Kaggle. The dataset contains Air Quality Index- AQI and air quality data in terms of daily and hourly levels of several stations across multiple cities in India, such as Aizawl, Amristar, Ahmedabad, Amaravati, Bhopal, Brajrajnagar, Bengaluru, Chennai, Delhi, Chandigarh, Coimbatore, Gurugram, Ernakulam, Hyderabad, Guwahati, Jorapokhar, Kolkata, Jaipur, Kochi, Mumbai, Lucknow, Talcher, Patna, Visakhapatnam, Shillong, and Thiruvananthapuram. Dataset Link: `https://www.kaggle.com/datasets/rohanrao/air-quality-data-in-india`.

4.2 Evaluation Metrics

The metrics that are used to evaluate the efficiency and error rate of the proposed model are stated as follows;

RMSE- The term RMSE is denoted as root mean square error, which is frequently utilized for measuring the differences between the predicted value of the model (or) the observed values and estimator.

$$RMSE = \frac{1}{k}\sum_{n=1}^{k}\left(\frac{(A_n - F_n)}{A_n}\right)^2 \quad (13)$$

Where n represents the variable and k indicates the counts of non-missing data points. A_n is the actual value, and F_n is the predicted value.

MAE- Mean Absolute Error (MAE) calculates the errors between paired observations expressing a similar phenomenon.

$$MAE = \frac{1}{k}\sum_{n=1}^{k}|(A_n - F_n)| \quad (14)$$

Where k indicates the total data point counts

MAPE- Mean Absolute Percentage Error is termed MAPE, which is one of the commonly used KPIs for evaluating the accuracy range of the forecast. MAPE is the addition of the individual absolute errors divided by the request.

$$MAPE = \frac{1}{k}\sum_{n=1}^{k}\left|\frac{(A_n - F_n)}{A_n}\right| * 100 \quad (15)$$

Where k represents the total times of the summation.

MSE- Mean Square Error is the average square difference between the actual and estimated values. The MSE values close to zero are considered better values

$$MSE = \frac{1}{k}\sum_{n=1}^{k}|(A_n - F_n)|^2 \quad (16)$$

Where k denotes the number of data.

4.3 Performance analysis

The proposed regression model error rate is analyzed in terms of the considered performance metric, and the obtained results are illustrated in Figure 5.

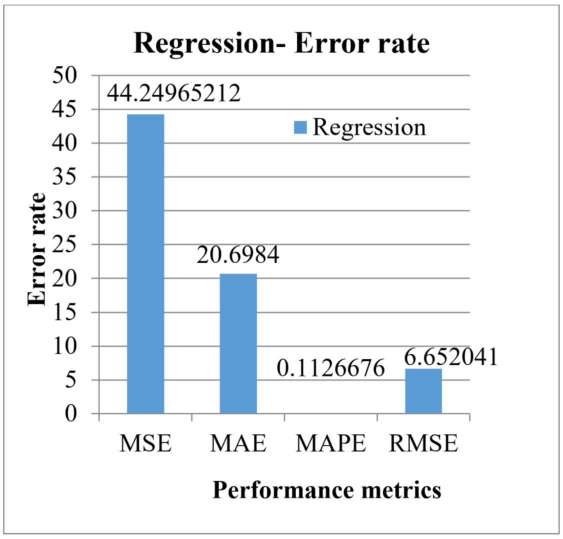

Fig.5: *Performance analysis of the proposed model.*

Table 5: *Performance analysis of the proposed model in terms of error rate.*

Performance analysis of Proposed model concerning an Error rate			
MSE	MAE	MAPE	RMSE
44.24965212	20.6984	0.1126676	6.652041

The performance analysis of the proposed model in terms of MSE, MAE, MAPE, and RMSE is shown in Figure 4 and Table 5. The proposed model showed 44.24 MSE, 20.69 MAE, 0.1126 MAPE and 6.652 RMSE. The MAPE error rate was low than the other error rate.

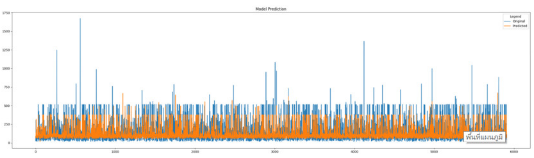

Fig.6: *Model Prediction.*

The prediction result of the proposed model is shown in Figure 6. The original values are denoted in "blue," and the predicated results are denoted in "Orange." The proposed model showed a better prediction rate with an error rate of 44.24 MSE, 20.69 MAE, 0.1126 MAPE, and 6.652 RMSE.

4.4 Comparative analysis

To prove the efficiency and effectiveness of the proposed regression model, the obtained error rate was compared with the existing state-of-the-art methods. Figures 7 and 8 show the comparative results of the proposed model with various existing regression models.

Table 6: *Comparative analysis between the proposed regression model and other regression models.*

Model	**MAE**	**RMSE**
Linear regression	29.227	38.49
Lasso regression	27.908	36.79
Ridge regression	27.907	36.79
Support vector regression	29.828	41.33
Proposed Regression	20.6984	6.65204

The proposed regression model is compared with the existing model [17], the comparison was made in terms of MAE and RMSE, and the obtained results are shown in Figure 5. From this observation, while comparing all existing models built with regression algorithms, the proposed model has lower MAE and RMSE. The MAE value of the proposed model is 20.69, and the RMSE range is 6.652, which is lower than the other methods, such as linear regression, Lasso regression, ridge regression, and support vector regression. Thus, the proposed regression model performs well in air pollution forecasting by utilizing Air Quality Data in India (2015 - 2020).

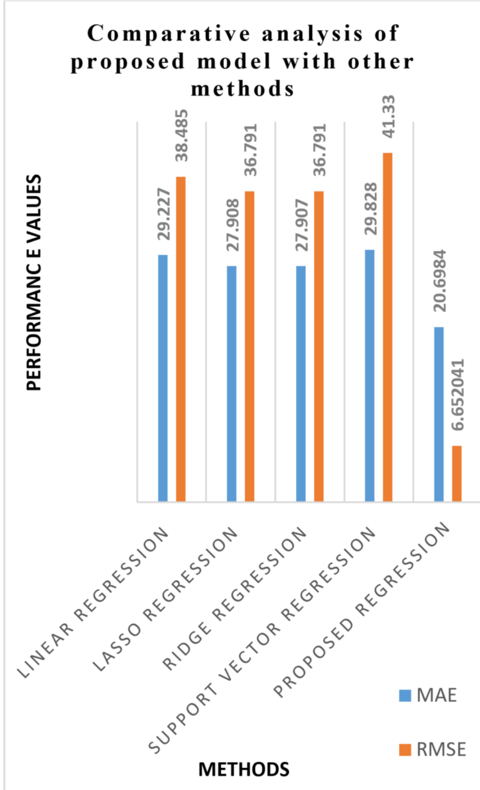

Fig.7: *Comparative analysis of the proposed regression with existing regression models [15].*

Table 7: *Comparative analysis between the proposed regression model and other regression models.*

Performance metrics	**LSTM**	**Hierarchical-attention based BiLSTM**
MSE	5147	44.24965212
MAE	7841	20.6984
MAPE	2452	0.1126676
RMSE	1457	6.652041

From the above table 7, it is revealed that the proposed Hierarchical-attention based BiLSTM performs better than the LSTM concerning evaluation metrics like MAE, MSE, RMSE, and MAPE.

The proposed model has several advantages over the existing models, like long-term dependencies, handling missing data, enhanced accuracy, and attention mechanism. Regarding evaluation metrics like RMSE, MAPE, MAE, and MSE, the proposed technique exhibits higher accuracy compared to the existing models in terms of time series prediction.

The comparison of the proposed regression model is made with the existing system [18] in terms of RMSE, MAPE, and MAE are shown in Figure 8. Therefore, by observing Figure 8, PLS showed 70.59

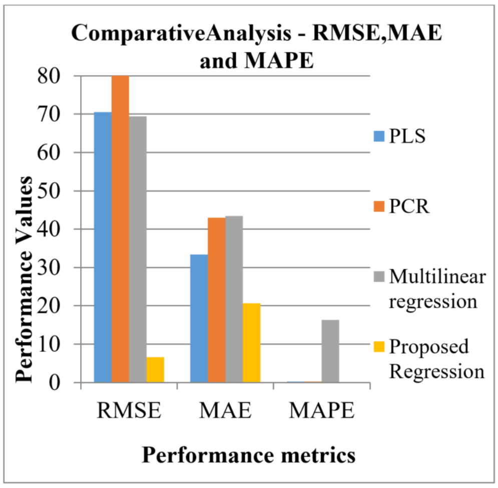

Fig.8: *Comparative analysis of proposed regression with existing model [16] in terms of RMSE, MAE, and MAPE.*

RMSE, 33.39 MAE, 0.2 MAPE, and PCR showed 91.26 RMSE, 43.01 MAE, and 0.25 MAPE. Then, the multi-linear regression showed 69.42 RMSE, 43.44 MAE, and 16.3 MAPE. Followed by this, the proposed model showed 6.6520 RMSE, 20.69 MAE, and 0.1126 MAPE. Where RMSE, MAE, and MAPE rate in the proposed model has been resulted as lower than the existing methods.

Hence, the results obtained from the proposed model regarding RMSE, MAE, MAPE, and MSE. The proposed regression model showed lower RMSE, MAE, MAPE, and MSE rates. This proves that the proposed regression performs well in predicting air pollution.

5. CONCLUSIONS

The study elucidated the time series prediction with a hybrid model of GAN with hierarchical-attention-based BiLSTM for an accurate prediction of air forecasting where the generative adversarial networks aided in imputing the missing value of data in the air quality dataset. The GAN imputation algorithm proposed in the present study assigned the missing value to the dataset with appropriate generator and discriminator optimization. The hierarchical-attention-based BiLSTM with the imputed data efficiently trained the model, and the hyperparameters maximized the model's performance, minimizing the predefined loss function that produced better prediction accuracy with minimal errors. The performance of the proposed hybrid model was assessed with evaluation metrics, namely RMSE, MSE, MAPE, and MAE. The experimental outcomes showed that the performance of the proposed model exhibited lower error values of 6.652041 (RMSE), 0.1126676 (MAPE), 20.6984 (MAE), and 44.24965212 (MSE). Thus the evaluation metrics showed the higher accuracy of the proposed model in time series prediction. The future scope of the proposed model could be an expansion of various use cases in the prediction of time series other than air quality.

References

[1] S. M. Hama, P. Kumar, R. M. Harrison, W. J. Bloss, M. Khare, S. Mishra, et al., "Four-year assessment of ambient particulate matter and trace gases in the Delhi-NCR region of India," *Sustainable Cities and Society*, vol. 54, p. 102003, 2020.

[2] M. Zamani Joharestani, C. Cao, X. Ni, B. Bashir, and S. Talebiesfandarani, "PM2. 5 predictions based on random forest, XGBoost, and deep learning using multisource remote sensing data," *Atmosphere*, vol. 10, p. 373, 2019.

[3] M. Dun, Z. Xu, Y. Chen, and L. Wu, "Short-term air quality prediction based on fractional grey linear regression and support vector machine," *Mathematical problems in engineering*, vol. 2020, 2020.

[4] J. Wang, H. Li, H. Yang, and Y. Wang, "Intelligent multivariable air-quality forecasting system based on feature selection and modified evolving interval type-2 quantum fuzzy neural network," *Environmental Pollution*, vol. 274, p. 116429, 2021.

[5] S. Agarwal, S. Sharma, R. Suresh, M. H. Rahman, S. Vranckx, B. Maiheu, *et al.*, "Air quality forecasting using artificial neural networks with real-time dynamic error correction in highly polluted regions," *Science of the Total Environment*, vol. 735, p. 139454, 2020.

[6] B. Liu, X. Tan, Y. Jin, W. Yu, and C. Li, "Application of RR-XGBoost combined model in data calibration of micro air quality detector," *Scientific Reports*, vol. 11, p. 15662, 2021.

[7] S. Siami-Namini, N. Tavakoli, and A. S. Namin, "A comparative analysis of forecasting financial time series using Arima, lstm, and bilstm," 2019., arXiv:1911.09512. [Online]. Available: `https://doi.org/10.48550/arXiv.1911.09512`

[8] S. Du, T. Li, Y. Yang, and S.-J. Horng, "Deep air quality forecasting using hybrid deep learning framework," *IEEE Transactions on Knowledge and Data Engineering*, vol. 33, pp. 2412-2424, 2019.

[9] Q. Tao, F. Liu, Y. Li, and D. Sidorov, "Air pollution forecasting using a deep learning model based on 1D convents and bidirectional GRU," *IEEE Access*, vol. 7, pp. 76690-76698, 2019.

[10] S. Pilleron, D. Sarfati, M. Janssen-Heijnen, J. Vignat, J. Ferlay, F. Bray, *et al.*, "Global cancer incidence in older adults, 2012 and 2035: a population-based study," *International journal of cancer*, vol. 144, pp. 49-58, 2019.

[11] Y. Han, J. C. Lam, V. O. Li, and Q. Zhang,

"A domain-specific Bayesian deep-learning approach for air pollution forecast," *IEEE Transactions on Big Data*, vol. 8, pp. 1034-1046, 2020.

[12] T. Kim, J. Kim, W. Yang, H. Lee, and J. Choo, "Missing Value Imputation of Time-Series Air-Quality Data via Deep Neural Networks," *International Journal of Environmental Research and Public Health*, vol. 18, p. 12213, 2021.

[13] S. Yang, M. Dong, Y. Wang and C. Xu, "Adversarial Recurrent Time Series Imputation," in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, no. 4, pp. 1639-1650, 2023.

[14] A. Bihlo, "A generative adversarial network approach to (ensemble) weather prediction," *Neural Networks*, vol. 139, pp. 1-16, 2021.

[15] Z. Zhao, J. Wu, F. Cai, S. Zhang, and Y.-G. Wang, "A hybrid deep learning framework for air quality prediction with spatial autocorrelation during the COVID-19 pandemic," *Scientific Reports*, vol. 13, p. 1015, 2023.

[16] K. H. Waseem, H. Mushtaq, F. Abid, A. M. Abu-Mahfouz, A. Shaikh, M. Turan, et al., "Forecasting of Air Quality Using an Optimized Recurrent Neural Network," *Processes*, vol. 10, p. 2117, 2022.

[17] K. Saikiran, G. Lithesh, B. Srinivas, and S. Ashok, "Prediction of Air Quality Index Using Supervised Machine Learning Algorithms," in *2021 2nd International Conference on Advances in Computing, Communication, Embedded and Secure Systems (ACCESS)*, pp. 1-4, 2021.

[18] M. Londhe, "Data mining and machine learning approach for air quality index prediction," *International Journal of Engineering and Applied Physics*, vol. 1, pp. 136-153, 2021.

Rupa Rajakumari R Peter is an Assistant Professor of Computer Science in Hislop College, Nagpur, India. Her Qualification is M.Sc. and M.Phil in Computer Science. And she also qualified the National Eligibility Test (NET). She Pursuing a Ph.D. at RTM Nagpur University, Nagpur, India. Possesses approximately 15 years of teaching experience, including 4 years at the undergraduate level. Specializes in the domains of Data Science which involves collecting, analysing, and interpreting large and complex datasets to extract meaningful insights and inform decision-making. It encompasses data mining, data visualization, and statistical analysis. Her achievement published a book in the field of Computer Science. Presented a paper at a National Level Conference.

Ujwal Ambadas Lanjewar is a Principal of Smt. Binzani Mahila Mahavidyalaya, Nagpur, India. His Educational Qualifications is Ph.D., MCA, M.Sc. (Stat.), MBA. He is 20 years of extensive teaching and research experience. And Engaged with graduate, postgraduate, and doctoral students. His achievements in Post-Doctoral Research Work submission in progress for Doctor of Science degree at RTM Nagpur University. Research Guide for five universities across Computer Science, Business Management, Applications, and Statistics. Mentored 15 students to Doctoral Degrees. He also Published 90 research papers in esteemed International Peer Reviewed Journals. Presented more than 20 research article in national and international conferences. He is authored seven books that contribute significantly to his research areas. Contribution to advisory committees of National and International Conferences. Former coordinator of the Special Task Force for the Board of Computer Application at RTM Nagpur University, Nagpur.