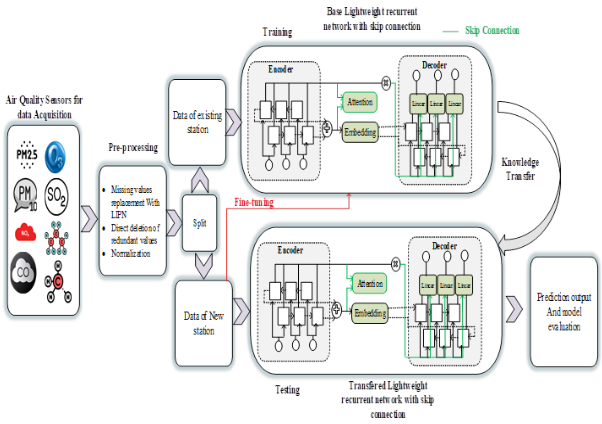
Global NEST Journal, Vol 26, No 5, 06096

An intelligent air quality monitoring system using quality indicators and transfer learning based lightweight recurrent network with skip connection

Periasamy S., Subramanian P*. and Surendran R.
Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, 602105, India
Received: 23/04/2024, Accepted: 08/05/2024, Available online: 15/05/2024
*to whom all correspondence should be addressed: e-mail: subramanianp.sse@saveetha.com
https://doi.org/10.30955/gnj.006096

Graphical abstract



Abstract

Rapid industrialization and urbanization have resulted in poor air quality, which poses a risk to human health by causing a variety of lung diseases. The precise forecast of air quality is of practical importance. Consequently, the development of an automated air pollution monitoring system based on environmental toxicology is required. Although advanced machine learning approaches can yield reasonable results in air quality prediction, they require more historical data collection. In order to address this problem, a lightweight recurrent network based on transfer learning with skip connection (LRN-SC) is proposed for air quality prediction. LRN-SC pretrains the model using data from an available station. The features that were learned from the previous station are retained, and the pre-trained model is then adjusted to fit the new one. After that, Transfer learning-based light weight recurrent network with skip connection (TL2RN-SC) is trained, and the model is tested using data from the new station. The proposed model reduces the decoding burden by adding skip contacts between the decoder and the linear forecasting layer. The simulation results show that the proposed model outperforms the existing models by attaining average RMSE and MAE of 0.974 and 2.63 respectively.

Keywords: Air quality; deep learning; recurrent network; simple recurrent network; and transfer learning

1. Introduction

Environmental monitoring systems evaluate the current state of air, soil and water, identify patterns, predict future conditions, send early warnings of potential hazards, and assist decision makers in sustainable development and environmental preservation. Chapman J *et al*. (2020) introduced environmental monitoring has evolved as smart environmental monitoring systems that incorporate modern sensors, Internet of Things (IoT) technologies, and machine learning (ML) techniques. Surendran R *et al*. (2021) developed a systems include various environmental monitoring applications, such as water and air monitoring. Hojjati-Najafabadi *et al*. (2022) implemented a water monitoring assesses the effects of contaminants and dangerous materials in lakes, rivers, and oceans, while air monitoring monitors air quality. Liu *et al*. (2020) executed a air quality forecasting becoming a popular research topic in the domains of pollution control, urban area development, and sustainable smart environmental design, specifically in rapidly developing countries like India. Lu *et al*. (2021) established the growing use of industrial technology and the growth of the transportation sector contribute to an increase in urban air pollution. Goodsite *et al*. (2021) introduced PM2.5 or tiny particles with a diameter of 2.5 microns or less in the atmosphere are the main contributor to air pollution. Human health will be directly affected by an increase in PM2.5 concentration by Wang *et al* (2021). Al-Janabi *et al*. (2021) added to this, there are several other toxins that adversely affect both the atmosphere and human health. Jion *et al*. (2021) stated the most common pollutants are sulfur dioxide (SO_2), ozone (O3), carbon monoxide (CO), and nitrogen oxides (NOx). Surendran R *et* *al*. (2023) describe the real-time forecasts of these factors have many practical and societal benefits. These data are essential for promoting sustainable development and play an important role in atmospheric management to reduce air pollution.

Castelli *et al*. (2020) introduce the technologies for predicting air quality use a variety of techniques, from sophisticated machine learning and deterministic models to traditional statistical techniques. Ma *et al*. (2021) determined a deterministic approaches are dependent on the mass transfer or thermodynamic hypothesis for in-depth modelling and analytical examination. It should be noted that the deterministic models that were realized using numerical simulation offer crucial information on self-governing variables and insightful commentary on system interactions during the design stage. Surendran R *et al*. (2023) and Khan *et al*. (2020) designed a the statistical forecasting approach uses historical time series data and quantitative analysis to forecast air quality. Regression analysis and time series analysis are two statistical prediction models that are often used by Masood *et al*. (2021). Jin *et al*. (2021) discussed about Machine learning (ML) approaches a core component of artificial intelligence (AI) and offers great opportunities to identify, classify, and forecast air quality indicators. Loy-Benitez *et al*. (2020) and Eslami *et al*. (2020) stated a deep learning methods such as long short-term memory (LSTM), gated recurrent unit (GRU), and convolutional neural network (CNN) have the ability to capture the non-linear characteristics of the air quality variables. Haq *et al*. (2022) implement a Long-term memory maintenance is a challenge for LSTM and GRU algorithms when predicting time series, particularly for long sequences. Mao *et al*. (2022) added a the static sequential input–output structure increases computation costs and slows the training rate of conventional RNNs such as LSTM. On the other hand, the recurrent computation is simplified by the Simple Recurrent Unit (SRU) through the avoidance of complicated gating mechanisms. As a result, SRU has faster training times and a lower demand for computer resources.

Ke *et al*. (2020) discuused a deep learning and advanced machine learning methods can yield accurate air and water quality predictions, their use requires a sufficiently large historical data collection. In the absence of sufficient historical data, neural networks perform poorly because they cannot identify the patterns concealed in the time series. For example, recently constructed air monitoring stations or those with insufficient data cannot supply enough samples for deep learning model training. Therefore, it is important to investigate a prediction model that may address or reduce the problem of data scarcity in recently constructed stations. The current recurrent models can handle the time-series data on air quality, but they cannot deal with the model's intrinsic sensitivity to outliers, which makes the model fit poorly for the peak value of quality indicators. Moreover, the network will become more complex due to the excessive parameter usage of current models. To this end, this paper proposes new lightweight deep learning models for air quality prediction. The main contributions of this research work are as follows:

To circumvent the issue of insufficient data in the air quality prediction model by introducing a transfer learning-based lightweight recurrent network with skip connection. To simplify recurrent computation by avoiding complicated gating techniques using SRU. Compared to a traditional RNN, it can result in shorter training times and less computing overhead. To address the issue of network learning deterioration and reduce gradient vanishing by adding a skip connection to the BiSRU decoder.

2. Related works

Examining air pollution is regarded as a crucial study in mapping the degree of pollution in various places. Machine learning techniques have extensive applications in the prediction, forecasting, and control of pollution levels. An automated technique of forecasting air quality based on machine learning predicts the levels of the main pollutants (PM2.5, PM10, SO2, NO2, O3, and CO) and their respective concentrations. Maltare *et al*. (2023) examined different machine learning techniques, including Seasonal Auto-Regressive Integrated Moving Average (SARIMA), Support vector machine (SVM) and LSTM to estimate the air quality index in India. Various pre-processing techniques, such as feature selection, outlier processing, and handling missing values, are employed in this study to manipulate data before feeding them to machine learning models. Kumar *et al*. (2023) used correlation analysis to preprocess the data set and identify important features. Five machine learning models, namely k-nearest neighbors (KNN), Gaussian naive Bayes (GNB), SVM, Random Forest (RF), and XGBoost, are used to forecast air quality after the data imbalance issue was resolved through resampling. According to this analysis, the SVM model has the least accuracy, while the GNB achieved the maximum accuracy. Monitoring air quality involved determining the intricate correlations between a variety of environmental parameters, including pollution concentrations, temperature, humidity, and wind speed. Conventional ML algorithms might struggle for capturing these intricate correlations efficiently, particularly when working with complex and non-linear data.

Some stations provide extremely dynamic, nonlinear and highly random spatial temporal correlations in their air quality data. These spatial-temporal elements can be effectively captured by deep learning algorithms. Santhanaraj R. K *et al*. (2023) introduced an integrated LSTM network (LSTM-FC) to predict PM2.5 contamination using previous air quality information, climatic data, weather prediction data and the day of the week. The two parts of this predictive model are as follows: modelling the local discrepancy of PM2.5 contamination utilizing an LSTM-based temporal simulant; and capturing the spatial dependences among the PM2.5 contamination of the primary station and neighboring stations by means of a neural network-based spatial collaborative model. QIN *et al*. (2019) presented a new integrated approach using CNN and LSTM to predict the density of urban PM2.5. This model extracted features from the input data spontaneously. The output layer took into account the time dependence of pollutants using an LSTM network. Zhang, Z. *et al*. (2024) introduced a layered deep learning framework, namely DL-Air, to forecast air quality. Here, all spatial relations in the data were encoded using the encoder. Then the spatiotemporal relationship analysis LSTM (STAA-LSTM) determined the degree of relationship between the forecasted and detected spatiotemporal data. Furthermore, STAA-LSTM predicted the upcoming spatio-temporal correlations in the hidden space. Finally, these correlations were appropriately decoded using a decoder to obtain the real forecast. The author introduced a bidirectional gated recurrent unit (Bi-GRU) network model to estimate the degree of various pollutants in the air.

Ahmed *et al*. (2024) analysis shows that the Recurrent Neural Network (RNNs) including GRU, LSTM, and their variants is an excellent tool for time series prediction because they can express express nonlinear relations and handle multiple-dimensional data in a nonlinear manner. LSTM and GRU still have significant computational restrictions due to their reliance on the hidden layer output from the prior calculation. The aforementioned LSTMs, GRU and its variations are excellent at identifying temporal dependencies during short to moderate time periods, but they could have trouble identifying long-term dependences in air quality data, which can be impacted by seasonal patterns, meteorological variations, and other outside variables. In the literature, this issue has been mitigated by using an attention mechanism because it has the capacity to extract crucial information. However, they are still unable to adequately capture all long-term dependences. Also, it presents the complicated decoding process, gradient disappearance, and learning deterioration problem. To tackle these issues, we used a special internal structure of Bi-SRU with attention and skip connection in this research. This model processed every input stream autonomously to remove temporal dependence without disturbing other inputs. Also, it transforms an RNN into a network that can be somewhat parallelized without computing the previous hidden layer output. Although the Bi-SRU network with attention and skip connection can yield good results in air quality prediction, it needs an adequate amount of historical data for training. The suggested strategy presents a transfer learning model to address or minimize the issue of data scarcity in recently constructed stations.

3. The proposed model

In this work, a new air quality prediction model is proposed by introducing lightweight networks based on transfer learning with recurrent learners. Initially, the source data is obtained from air pollutant data sets. After that, preprocessing is applied to the collected data using the linear interpolation (LIPN) algorithm to fill missing values. The direct deletion method eliminates values that are unnecessary or redundant. After preprocessing, a novel lightweight network based on transfer learning is proposed for air quality prediction. This approach used advanced deep learning techniques with transfer learning to address the problem of data scarcity in newly constructed air quality monitoring stations. Initially, an LRN-SC is proposed for pre-training the model using data from an existing station. Subsequently, it froze the first few layers of the basic model and the data from a new station is used to modify the remaining hidden layers. In this case, the features learned from the prior station can be retained and the model can be adjusted to meet the current station. After that, TL2RN-SC is trained and the data of the new station is then used to test the model. The proposed model reduced the decoding burden by insertion of skip connections between the input of the decoder and the linear prediction layer. The complete architecture of the proposed air quality monitoring system is illustrated in Figure 1.

3.1. Data pre-processing

In this work, the data obtained from the Kaggle website is used to train the neural network. It is a time series of data of different Indian stations from 2017 to 2020. It includes hourly readings for a number of gases, including PM2.5, PM10, nitrogen oxides (NOx), carbon monoxide (CO), ozone (O3) and sulfur dioxide (SO2). The parameters obtained from various stations might have irregularities including missing or erroneous data. This may be the result of sensor problems or potential data storage errors. These irregularities will cause an extreme deviation between the predicted and actual values.

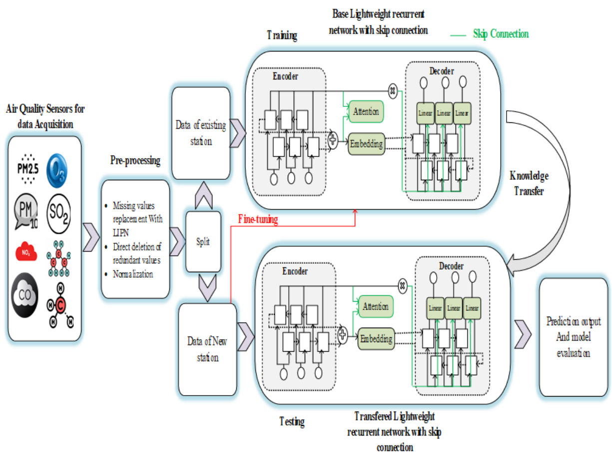


Figure 1. Proposed air quality monitoring framework

The air quality parameters acquired from the source is a time series data. Time series data are data that are collected over a predetermined period of time. The time series air quality parameters are defined as

$$X_{l,m} = \left(\left(s_{l,1}, Z_1\right), \ldots, \left(s_{l,m}, Z_m\right)\right) \tag{1}$$

Where $X_{l,m}$ represents the time series data of l–th air quality parameter whose length is m. Each parameter has a sampling interval of one day. This means that all parameters are measured every day simultaneously. When the value $S_{l,i}$ at Z_i is not present, the estimated values can be obtained using the LIPN approach. It is possible to construct the LIPN function as

$$S_{l,i} = s_{l,j} + \frac{s_{l,j} - s_{l,k}}{Z_j - Z_k}\left(Z_i - Z_j\right) \tag{2}$$

If there are some missing values in air quality parameters, the LIPN algorithm determines the two nearest instants Z_j and Z_k. After that the missing value at the instant $Z_i$is calculated using the values $S_{l,j}$ and $S_{l,k}$. Here $S_{l,j}$ represents the computed value for the missing data $S_{l,i}$. Also, the input data set was normalized using the Z-score standardization method to increase the accuracy of the prediction of the model and the training speed.

3.2. LRN-SC based on transfer learning

Although the advanced deep learning model can potentially achieve excellent performance for air quality prediction, its effectiveness is limited as there is inadequate training data. The network does not obtain sufficient knowledge due to a data shortage. To address this problem, a Transfer learning-based light weight recurrent network with skip connection (TL2RN-SC) is introduced in this paper. This strategy combined advanced transfer learning and recurrent learning techniques to solve the issue of data scarcity in recently constructed air quality monitoring stations. The architecture of TL2RN-SC is shown in Figure 2.

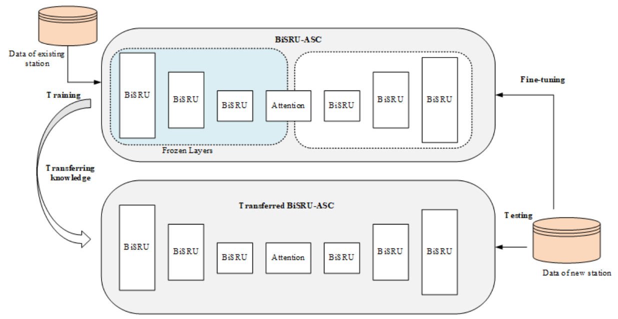


Figure 2. Architecture of TL2RN-SC

3.3. BiSRU-ASC

Traditional Bi-GRU models with recurrent learning struggle to maintain long-term memory for time-series prediction. In addition, expanding the application of deep learning in various domains is highly dependent on the development and improvement of lightweight models. In this paper, a new BiSRU-ASC is proposed as a base model for the proposed TL2RN-SC. The architecture. Assume that S = [S_1, S_2, ...S_n] Tis time-series data. The proposed BiSRU-ASC is made up of BiSRU based auto encoders. The BiSRU network consists of direction-dependent SRUs for extracting time dependencies from input time series data in forward and backward direction. The hidden layer output of the encoder is considered as the embedding vector H_evto create air indicator. The proposed model concatenates the hidden layer output and it is given as input to the attention layer along with the encoder output O_enc. The weight vector can be computed by the attention layer in various time steps. After that, the encoder output results are multiplied by the vector to produce the _out, which emphasizes the importance of the primary timesteps data. During the decoding phase, Å_outis supplied as an input vector to the BiSRU decoder. Also, the proposed model concatenates _out with the decoder output via skip link. The concatenated output is provided as input to the linear prediction layer to augment the feature information and relieve the decoding burden.

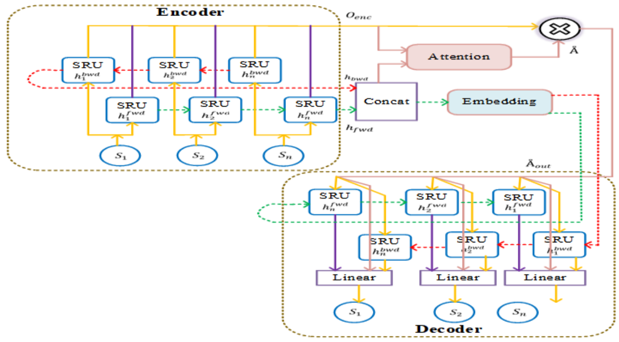


Figure 3. Architecture of BiSRU-ASC

3.4. BiSRU encoding

Traditional LSTM and GRU depend on the output h_(n-1) of the prior run. But the special interior structure of SRU allows one to process individual input steps self-sufficiently, remove time dependence during operation and transform an RNN into a network that may be partially parallelized without calculating h_(n-1). In Figure 4, the structures of the LSTM, GRU, and SRU models are compared.

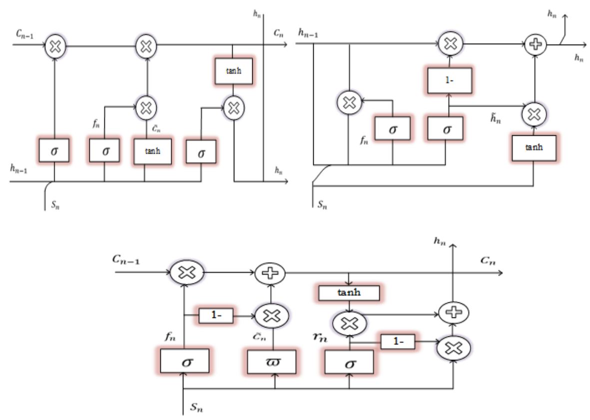


Figure 4. Recurrent networks (a) LSTM (b) GRU (c) SRU

In Figure 4, the two addition and multiplication variables are indicated as '+' and "×" respectively. Also, σ and *tanh* denote the weight matrix, sigmoid activation, and hyperbolic tangent activation correspondingly.

$$\sigma(s) = \frac{1}{1+e^{-s}} \tag{3}$$

$$tanh(s) = \frac{e^{s} - e^{-s}}{e^{s} + e^{-s}} \tag{4}$$

SRU receives S_n as present input and calculates the outputs without requiring the results from the previous run. The SRU data flow is shown in Figure 5.

$$f_n = \sigma\left(\varpi_f S_n + \beta_f\right) \tag{5}$$

where the forgetting gate f_n represents the level of historical information. After mapping the S_n linearly, the sigmoid activation function of f_n yields a value between 0 and 1.

$$\tilde{S}_n = \varpi \times S_n \tag{6}$$

Equation (6) multiplies S_n with the weight matrix ϖ for obtaining the $\tilde{S}_n$ and is given as input to the next cell C_n. The reset gate r_n is a sigmoid gate that regulates the amount of information update using the following expression.

$$r_n = \sigma\left(\varpi_r S_n + \beta_r\right) \tag{7}$$

Following the preceding action, data is still transferred to SRU to determine the output h_n Figure 5 shows the data flow of SRU. The cell status C_n is renovated as follows:

$$C_n = f_n \times c_{n-1} + \left(1 - f_n\right) \times \tilde{S}_n \tag{8}$$

In (8), the first term multiplies forgetting gate f_n with the previous cell status c_{n-1} for detecting the amount of overlook data that have been overlooked in the old status. The second term multiplies the remembered data with present input data $\tilde{S}_n$ for detecting the amount of reserved data. Finally, it adds the reserved data with the old cell status for the formation of new cell status C_n. After that, the present moment output h_n is detected as follows:

$$h_n = r_n \times tanh\left(c_n\right) + \left(1 - r_n\right) \times S_n \tag{9}$$

It multiplies the *tanh* function with the r_n after activating the new cell status C_n.

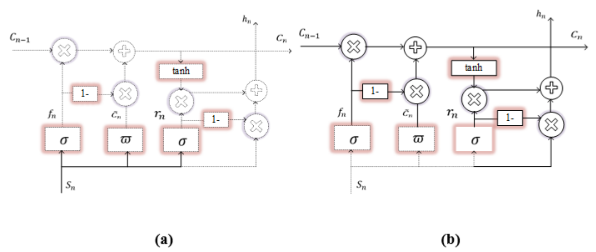


Figure 5. SRU data flow direction (a) Direction 1 (b) Direction 2

Given that $\ddot{h}$ is the hidden node of BiSRU, the encoder output and the last hidden layer output are obtained as follows:

$$\left(h_{enc}^{1\times 2\ddot{h}}, O_{enc}^{n\times 2\ddot{h}}\right) = f_{enc}\left(S^{n\times m}\right) \tag{10}$$

$$h_{enc}^{1\times 2\ddot{h}} = h_{fwd}^{1\times \ddot{h}} \oplus h_{bwd}^{1\times \ddot{h}} \tag{11}$$

where $f_{enc}\left(\square\right)$ denotes the fundamental function of the BiSRU encoder. The hidden forward and backward states are represented as h_{fwd}, and h_{bwd} individually. This h_{fwd} and h_{bwd} are concatenated to get encoder output h_{enc}.

3.5. Attention mechanism and BiSRU decoding

In this framework, an attention mechanism is utilized to differentiate the influence of various data from the timestep in the embedding vector. The attention layer used $h_{enc}^{1\times 2\ddot{h}}$ and $O_{enc}^{n\times 2\ddot{h}}$ to compute attention weight. It copies the initial size of $h_{enc}^{1\times 2\ddot{h}_n}$ times for aligning with the size of $O_{enc}^{n\times 2\ddot{h}}$ vector. Then, the attention layer receives $h_{enc}^{n\times 2\ddot{h}}$ and $O_{enc}^{n\times 2\ddot{h}}$ as input for computing the attention weight $\mathring{A}^{1\times n}$ at every time interval. After that, the proposed mechanism multiplies the $\mathring{A}^{1\times n}$ with $O_{enc}^{n\times 2\ddot{h}}$ for obtaining the $\mathring{A}_{out}^{1\times 2\ddot{h}}$ vector. The decoder unit of BiSRU receives $\mathring{A}_{out}$ as input to compute the output vector and hidden status as given below:

$$h_{enc}^{1\times 2\ddot{h}} \xrightarrow{copy} h_{enc}^{n\times 2\ddot{h}} \tag{12}$$

$$\mathring{A}^{1\times n} = Atten\left(h_{enc}^{n\times 2\ddot{h}} \oplus O_{enc}^{n\times 2\ddot{h}}\right) \tag{13}$$

$$\mathring{A}_{out}^{1\times 2\ddot{h}} = \mathring{A}^{1\times n} \times O_{enc}^{n\times 2\ddot{h}} \tag{14}$$

$$\left(h_{dec}^{1\times 2\ddot{h}}, O_{dec}^{1\times 2\ddot{h}}\right) = f_{dec}\left(\mathring{A}_{out}^{1\times 2\ddot{h}}, h_{enc}^{1\times 2\ddot{h}}\right) \tag{15}$$

After that, $\mathring{A}_{out}$ is combined with $O_{dec}$for giving the result as input to the linear prediction layer. It decodes S'_n , by repeating the above steps *n* times for predicting the complete encoded inputs $S' = \left[S'_1, S'_2, \ldots S'_n\right]^T$. The prediction error at moment *j* is $\varepsilon_j = S'_j - S_j$. The proposed TL^{2}RN-SC is trained for minimizing the prediction error using the following expression.

$$E = \frac{1}{2}\sum_{j=1}^{n}\left(\varepsilon_{j1}\right)^2 \tag{16}$$

where ε_{j1} denotes the 1-norm operation. It converges more quickly and yields a more reliable model compared to 2-norm operation. The last encoder hidden state h_{enc} can be thought of as the compact depiction of the input *S*. Suppose that the suggested framework contains several BiSRU layers, the final hidden states of all layers are combined to obtain the embedding vector as given below:

$$Z_n = h_{enc}^1 \oplus h_{enc}^2 \oplus \ldots h_{enc}^L \tag{17}$$

where h_{enc}^L denotes the *L*–th layer's last hidden state vector, Z_n denotes the embedding vector for the input time series data, and *L* is the total number of BiSRU layers.

4. Results and discussion

To verify the effectiveness of the proposed methodology, the air quality data is collected from https://www.kaggle.com/datasets/rohanrao/air-quality-data-in-india?select=station_day.csv. This data set contains sensor measurements of various pollutants, such as PM2.5, PM10, NO, NO2, NOx, NH3, CO, SO2 and O3. This section first examines the periodicity of the data in the data set using the violin plot. The violin plot group the data into various time periods through the combination of all the data across years and months. The violin plots of different contaminants over time are presented in Figure 6, both on an annual and monthly basis. India experiences a decrease in pollution between June and August. This may be due to the onset of a monsoon in the Indian subcontinent at this time. The smaller values for 2020 indicate a significant drop in pollution. During the COVID-19 pandemic, India implemented strict lockdowns that suspended all industrial and human activity. This is the reason for the significant drop in pollution.

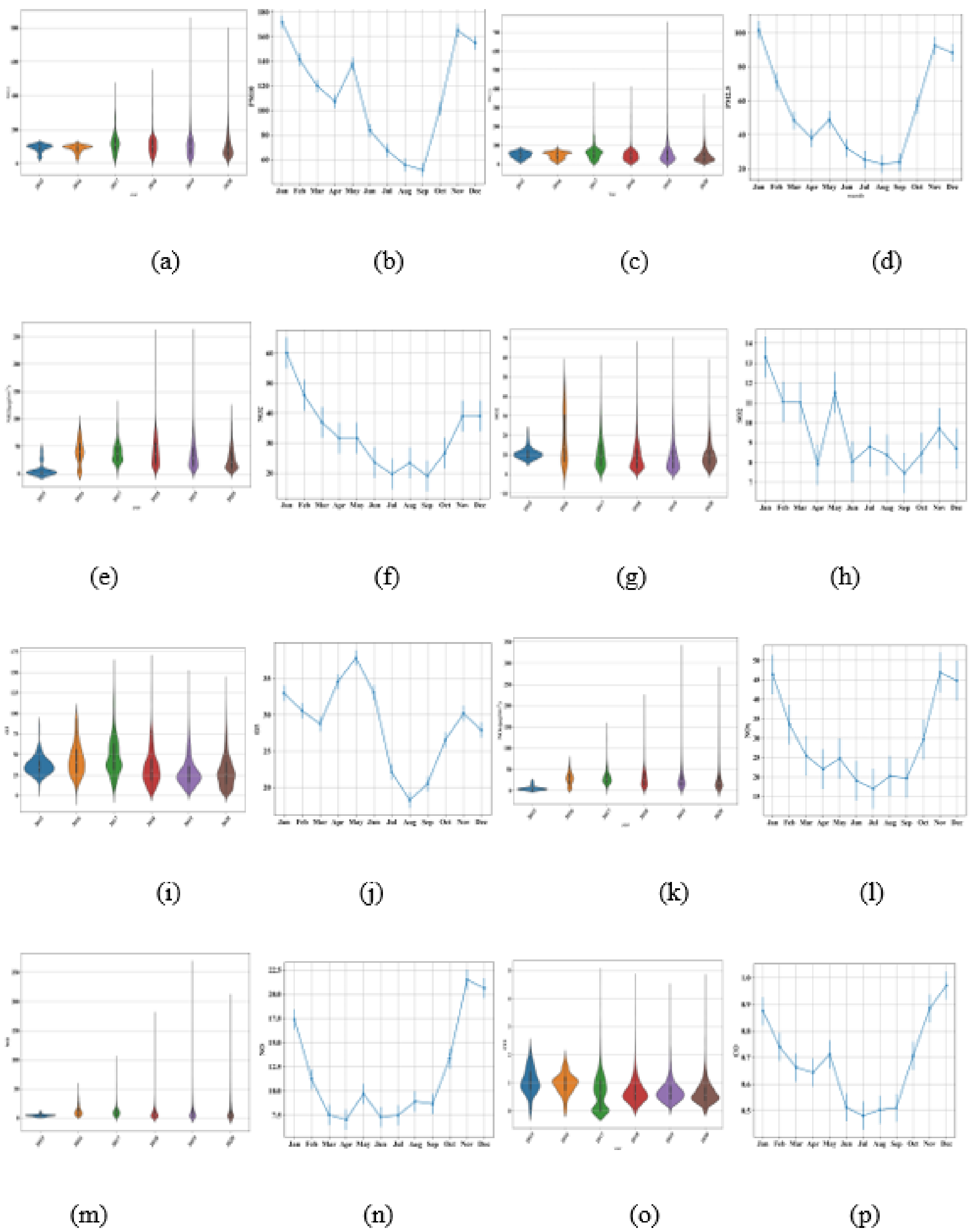


Figure 6. Distribution of the data (a) PM10 yearly (b) PM10 monthly (c) PM2.5 yearly (d) PM2.5 monthly (e) NO_2 yearly (f) NO_2 Monthly (g) SO_2 yearly (h) SO_2 monthly (i) O_3 yearly (j) O_3 monthly (k) NOx yearly (l) NOx monthly (m) NO yearly (n) NO monthly (o) CO yearly (p) CO monthly

After collecting the input data, they are utilized to construct the base LRN-SC model. The model parameters of the deep learning model must be determined to produce the best possible outcome. The hyperparameters of the proposed model are provided in Table 1.

Following the construction of time series samples and the determination of model parameters, the LRN-SC model is used to model data and forecast the level of air pollution for the coming hour. The prediction performance of the suggested model is assessed using three assessment indicators.

Table 1. Hyperparameters settings

Hyperparameters	Values
No of hidden layers	5
No of hidden nodes	30/50/75/100
Learning rate	0.001
Epoch	100
Optimizer	Mini-Batch Gradient Descent (MBGD)

$$RMSE = \sqrt{\frac{1}{m}\sum_{j=1}^{m}\left(o_j - o_j^*\right)^2} \quad (18)$$

$$MAPE = \frac{1}{m}\sum_{j=1}^{m}\frac{\left|o_j - o_j^*\right|}{o_j} \quad (19)$$

$$MAE = \frac{1}{m}\sum_{j=1}^{m}\left|o_j - o_j^*\right| \quad (20)$$

where m denotes the number of samples. The actual and predicted values of the j-th sample are represented as o_j and o_j* respectively. Reduced values of these three variables indicate improved model performance and increased forecast accuracy.

4.1. Performance analysis

The suggested TL²RN-SC model used the idea of transfer learning to improve prediction performance for stations with inadequate data. As a result, the source data is used to pretrain the model, while the target data are used to fine-tune and test the model. In this section, the effectiveness of the proposed transfer learning model is validated by comparing it with simple Bi-SRU, BiSRU-ASC and transfer learning based BiSRU-ASC (i.e TL²RN-SC) in Figure 7, Figure 8, Figure 9, Figure 10, Figure 11, Figure 12, Figure 13 and Figure 14. The findings demonstrate that the Bi-SRU and BiSRU-ASC models are unable to acquire sufficient information for prediction and produce large error rates as a result of insufficient training data. To address the issue of data scarcity, the proposed TL²RN-SC transfer knowledge gained from the pre-trained BiSRU-ASC to new stations. Thus, it overtakes all other base models. This shows the effectiveness of transfer learning in the proposed air quality prediction model.

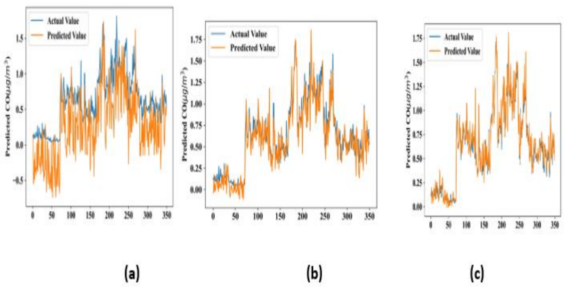


Figure 7. Comparative analysis of predicted and actual CO (a) Bi-SRU, (b) BiSRU-ASC, and (c) TL2RN-SC

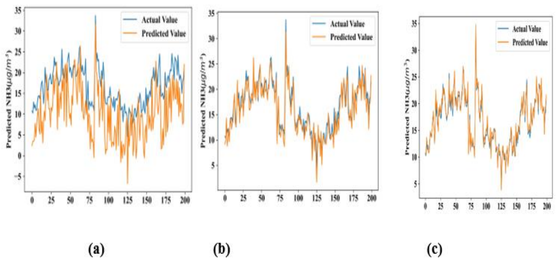


Figure 8. Comparative analysis of predicted and actual NH3(a) Bi-SRU (b) BiSRU-ASC and (c) TL²RN-SC

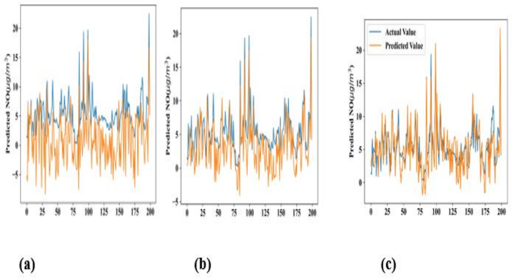


Figure 9. Comparative analysis of predicted and actual NO(a) Bi-SRU (b) BiSRU-ASC and (c) TL²RN-SC

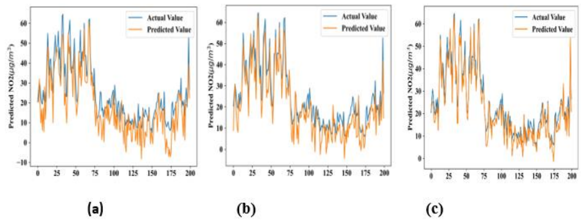


Figure 10. Comparative analysis of predicted and actual NO2(a) Bi-SRU (b) BiSRU-ASC and (c) TL²RN-SC

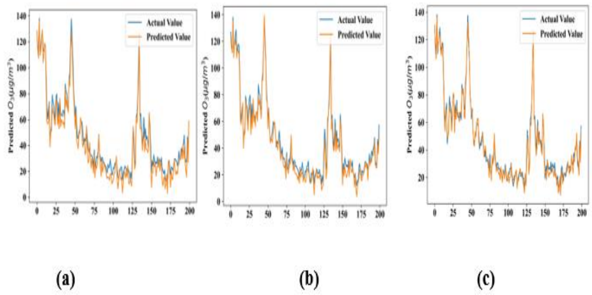


Figure 11. Comparative analysis of predicted and actual O3(a) Bi-SRU (b) BiSRU-ASC and (c) TL²RN-SC

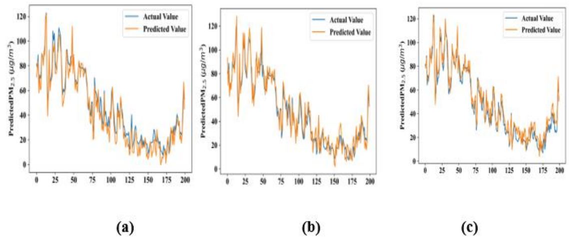


Figure 12. Comparative analysis of predicted and actual PM2.5(a) Bi-SRU (b) BiSRU-ASC and (c) TL²RN-SC

4.2. Comparative analysis

The performance of the proposed TL2RN-SC model for different pollutants including PM2.5, PM10, NO, NO2, NOx, NH3, CO, SO2 and O3 are listed in Table 2. It is obvious that CO, PM10 and PM2.5 have the lowest indicator values.

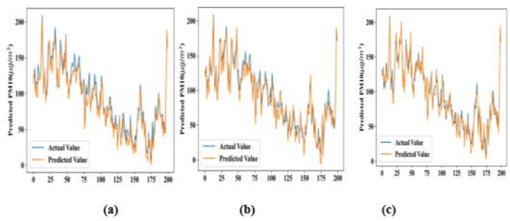


Figure 13. Comparative analysis of predicted and actual PM10(a) Bi-SRU (b) BiSRU-ASC and (c) TL2RN-SC

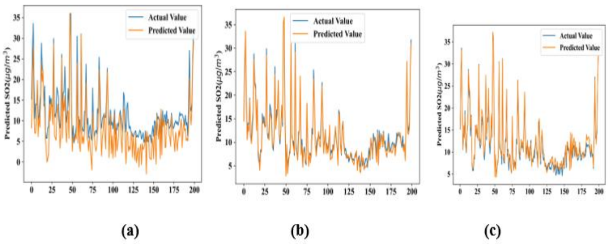


Figure 14. Comparative analysis of predicted and actual SO2 (a) Bi-SRU (b) BiSRU-ASC and (c) TL2RN-SC

This study compares the TL2RN-SC model with four additional deep learning techniques, such as the stacked BiLSTM model (S-BiLSTM), the stacked BiGRU (S-BiGRU), the Bi-SRU and the BiSRU-ASC in order to further test the performance of the model. In this instance, 30% of the samples from the station Secretariat, Amaravati, are utilized for validation, and 70% are used for training. The results of the comparison are shown in Table 3. Stacked BiLSTM models are susceptible to overfitting, especially when working with noisy or sparse data sets. As a result, S-BiLSTM performs worse than S-BiGRU. Unlike BiSRU models, S-BiGRU might have some drawbacks in terms of complexity, training effectiveness, and interpretability. However, the suggested BiSRU-ASC includes an attention mechanism that may enhance performance through the extraction of more insightful features. In addition, it allows the model to concentrate on pertinent segments of the input series.

Table 2. Comparative analysis of TL^{2}RN-SC model for different pollutants

Pollutants	RMSE	MAE	MAPE
PM2.5	0.972	2.14	8.36
PM10	0.968	2.75	8.99
NO	0.978	2.83	8.23
NO_2	0.987	2.75	8.54
NOx	0.973	2.77	8.71
NH_3	0.978	2.53	8.19
CO	0.961	2.39	8.37
SO_2	0.976	2.84	8.43
O_3	0.979	2.67	8.93

Table 3. RMSE performance of different models

Models	PM2.5	PM10	NO	NO_2	NOx	NH_3	CO	SO_2	O_3
S-BiLSTM	0.872	0.868	0.852	0.861	0.863	0.879	0.851	0.862	0.855
S-BiGRU	0.896	0.884	0.873	0.882	0.894	0.902	0.872	0.889	0.871
Bi-SRU	0.912	0.928	0.908	0.913	0.917	0.925	0.891	0.906	0.911
BiSRU-ASC	0.954	0.942	0.923	0.957	0.941	0.959	0.923	0.934	0.949
TL^{2}RN-SC	0.972	0.968	0.978	0.987	0.973	0.978	0.961	0.976	0.979

Table 4. RMSE Comparative analysis with other reported works

Model	Air pollutants						
	PM2.5	PM10	SO_2	NO_2	CO	NH_3	O_3
LSTM-FC	0.835	0.841	0.839	0.840	0.843	0.838	0.835
2DCNN-LSTM	0.799	0.804	0.800	0.828	0.807	0.802	0.804
DL-Air	0.947	0.948	0.947	0.961	0.957	0.950	0.953
Bi-GRU	0.786	0.874	0.821	0.785	0.882	-	-
Proposed	0.972	0.968	0.976	0.987	0.961	0.978	0.979

In addition, the decoding ability of the decoder is enhanced by using skip connections between the decoder input and the linear prediction layer. Furthermore, the TL2RN-SC model addresses the problem of data scarcity by transferring knowledge from a known station to a new station. As a result, it performs better than all other standalone models. The R2 values of the suggested air quality parameter prediction are compared with some existing methods, including LSTM-FC, 2DCNN-LSTM, DL-Air, and Bi-GRU. Although long-range dependencies can be captured by LSTM networks with sequential data, the fully connected layers in LSTM-FC might not fully exploit this feature, which could result in suboptimal performance in applications that need long-range dependencies modelling. Furthermore, the complicated structure of 2D CNN-LSTM models might make them difficult to train and susceptible to overfitting. Bi-GRU models struggle to maintain long-term memory for time series prediction. Moreover, overfitting may occur more frequently in DL-Air due to the increasing complexity of STAA-LSTMs. Table 4 indicates that the suggested TL2RN-SC has greater R2 values compared to other methods. It demonstrates an R2 improvement of more than 2.7 % over the baseline method with the highest performance (DL-Air). The enhanced R2 values imply the exactness of the proposed model in the prediction of air quality. The better performance of TL2RN-SC demonstrates its effectiveness in precisely capturing the temporal relationship and its impacts on the predicted values with simple architectures.

5. Conclusions

The changing nature of the environment, the unpredictability of pollutants, and spatial and temporal variability make the prediction of air quality a difficult task. The current study examines air pollution data for six years from different Indian cities. Furthermore, a transfer learning-based lightweight recurrent network with skip connection (TL2RN-SC) model is shown to enhance air pollution forecasting precision, particularly for new monitoring stations with a limited amount of historical data. In TL^{2}RN-SC, the SRU model was used with attention mechanism and a skip connection to minimize the decoding burden and enhance the significant feature extraction ability. Experimental results showed that the TL2RN-SC method has the highest model validation accuracy in terms of RMSE (0.974) and MAE (2.63). In future, the proposed work will extend with higher number of parameters and larger datasets.

References

Ahmed A.A.M., Jui S.J.J., Sharma E., Ahmed M.H., Raj N. and Bose A. (2024). An advanced deep learning predictive model for air quality index forecasting with remote satellite-derived hydro-climatological variables. Science of The Total Environment, 906, 167234.

Al-Janabi S., Alkaim A., Al-Janabi E., Aljeboree A. and Mustafa M. (2021). Intelligent forecaster of concentrations (PM2. 5, PM10, NO2, CO, O3, SO2) caused air pollution (IFCsAP). Neural Computing and Applications, 21, 14199–14229.

Castelli M., Clemente F.M., Popovic A., Silva S. and Vanneschi L. (2020). A machine learning approach to predict air quality in California. Complexity, 1–21.

Chapman J., Vi Khanh Truong., Elbourne A., Gangadoo S., Cheeseman S., Rajapaksha P., Latham K., Russell J., Crawford. and Cozzolino D. (2020). Combining chemometrics and sensors: Toward new applications in monitoring and environmental analysis. Chemical Reviews, 120, 6048–6069.

Dongming Q., Yu J., Zou G., Yong R., Zhao Q. and Zhang B. (2019). A novel combined prediction scheme based on CNN and LSTM for urban PM 2.5 concentration. Ieee Access, 7, 20050–20059.

Eslami E., Yunsoo C., Lops Y. and Sayeed A. (2020). A real-time hourly ozone prediction system using deep convolutional neural network. Neural Computing and Applications, 32, 8783–8797.

Goodsite M.E., Hertel O., Johnson M.S. and Jørgensen N.R. (2021). Urban air quality: sources and concentrations. Air Pollution Sources, Statistics and Health Effects, 193–214.

Haq K.R.A. and Harigovindan V.P. (2022). Water quality prediction for smart aquaculture using hybrid deep learning models. Ieee Access, 10, 60078–60098.

Hojjati-Najafabadi., Akbar., Mojtaba Mansoorianfar., Tongxiang Liang., Khashayar Shahin. and Hassan Karimi-Maleh. (2022). A review on magnetic sensors for monitoring of hazardous pollutants in water resources. Science of The Total Environment, 824, 153819–153844.

Jin N., Zeng Y., Yan K. and Ji Z. (2021). Multivariate air quality forecasting with nested long short term memory neural network. IEEE Transactions on Industrial Informatics, 17, 8514–8522.

Jion M.M.M.F., Jannatun N.J., Mia M.Y., Ali M.A., Islam M.D.S., Ibrahim S.M., Pal S.C. *et al*. (2023). A critical review and prospect of NO2 and SO2 pollution over Asia: Hotspots, trends, and sources. Science of the Total Environment, 876, 162823–162851.

Ke H., Gong S., He J., Zhang L., Cui B., Wang Y., Mo J., Zhou Y. and Zhang H. (2022). Development and application of an automated air quality forecasting system based on machine learning. Science of The Total Environment, 806, 151189–151204.

Khan F.M., and Gupta R. (2020). ARIMA and NAR based prediction model for time series analysis of COVID-19 cases in India. Journal of Safety Science and Resilience, 1, 12–18.

Kumar K. and Pande B.P. (2023). Air pollution prediction with machine learning: a case study of Indian cities. International Journal of Environmental Science and Technology, 20, 5333–5348.

Liu X., Jayaratne R, Thai P., Kuhn T., Zing I., Christensen B., Lamont R. *et al*. (2020). Low-cost sensors as an alternative for long-term air quality monitoring. Environmental research, 184, 1094001–109438.

Loy-Benitez J., Heo S. and Yoo C. (2020). Soft sensor validation for monitoring and resilient control of sequential subway indoor air quality through memory-gated recurrent neural networks-based autoencoders. Control Engineering Practice, 97, 104110–104330.

Lu J., Li.B., Li H. and Abdo Al-Barakani. (2021). Expansion of city scale, traffic modes, traffic congestion, and air pollution. Cities, 108, 102974–102994.

Ma N., Aviv D., Guo H. and Braham W.W. (2021). Measuring the right factors: A review of variables and models for thermal comfort and indoor air quality. Renewable and Sustainable Energy Reviews, 135, 110436–110452.

Maltare N.N. and Vahora S. (2023). Air Quality Index prediction using machine learning for Ahmedabad city. Digital Chemical Engineering, 100077–100093.

Mao S. and Sejdic E. (2022). A review of recurrent neural network-based methods in computational physiology. IEEE transactions on neural networks and learning systems, 34, 6983–7003.

Masood A., and Ahmad K. (2021). A review on emerging artificial intelligence (AI) techniques for air pollution forecasting: Fundamentals, application and performance. Journal of Cleaner Production, 322, 129045–129072.

Santhanaraj R.K., Rajendran S., Romero C.A.T. and Murugaraj S.S. (2023). Internet of Things Enabled Energy Aware Metaheuristic Clustering for Real Time Disaster Management. Comput. Syst. Sci. Eng., 45, 1561–1576.

Surendran R., Alotaibi Y. and Subahi A.F. (2023), Lens-Oppositional Wild Geese Optimization Based Clustering Scheme for Wireless Sensor Networks Assists Real Time Disaster Management. Comput. Syst. Sci. Eng., 46, 835–851.

Surendran R., Alotaibi Y. and Subahi A.F. (2023). Wind Speed Prediction Using Chicken Swarm Optimization with Deep Learning Model. Computer Systems Science & Engineering, 46, 3.

Surendran R., Tamilvizhi T. and Lakshmi S. (2021). Integrating the Meteorological Data into a Smart City Service Using Cloud of Things (CoT). In Emerging Technologies in Computing: 4th EAI/IAER International Conference, iCETiC 2021, Virtual Event, August 18–19, 2021, Springer International Publishing, 4, 94–111.

Wang L., Luo D., Liu X., Zhu J., Wang F., Li B. and Li L. (2021). Effects of PM2. 5 exposure on reproductive system and its mechanisms. Chemosphere, 264, 128416–128436.

Zhang Z., Zhang S., Chen C. and Yuan J. (2024). A systematic survey of air quality prediction based on deep learning. Alexandria Engineering Journal, 93, 128–141.