Multimedia Tools and Applications (2022) 81:6031–6049
https://doi.org/10.1007/s11042-021-11734-x

Analysis of deep learning approaches for air pollution prediction

Veena Gugnani[1] · Rajeev Kumar Singh[1]

Received: 25 April 2021 / Revised: 24 August 2021 / Accepted: 5 November 2021 /
Published online: 5 January 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract
Due to the urban and industrial growth, many evolving countries suffer from excessive air pollution. The growing concern about air pollution has been raised by the government and people because it affects individual's health and sustainable development globally. Recent methods for the prediction of air quality primarily use vast models; furthermore, these approaches yield inconsistent results, inspiring us to inspect air quality prediction methods based on deep learning architectures. While there is a range of efforts in the literature to figure pollution levels, recent developments in deep learning techniques, along with the incorporation of more data, offer more precise predictive accuracy. The paper analyses the previous deep learning frameworks proposed for air quality prediction. This paper discusses and reviews the different deep learning architectures with their advantages and disadvantages for air pollution forecasting.

Keywords Deep learning · Air pollution · LSTM · Particulate matter · Spatiotemporal deep learning

1 Introduction

In today's scenario, several cities are experiencing air pollution issues due to environmental and climate change. Corresponding to the latest study, disclosure to Particulate Matter (PM), a form of air pollution, causes around 4.2 million death and rates 5th amongst all risks worldwide [22]. This drastic growth motivates us to concentrate on the issue of air pollution and suggest a resolution on the basis of IoT data. Air pollution(AP) causes respiratory complications, hospitalization for heart or lung conditions, premature death, and various impacts on communities [9]. The concurrent access of air quality information and taking preventive measures and timely safety is vital to stop health threats brought on via Air Pollution exposure [56]. Hence, the monitoring of air quality is quite essential.

✉ Veena Gugnani
veenaaneja@gmail.com

Rajeev Kumar Singh
rajeev.mits1@gmail.com

[1] Madhav Institute of Technology and Science, Gwalior, Madhya Pradesh, India

Small particles of solid or liquid disperse in gasoline are generally referred to as Particulate Matters. Meanwhile, particles and fuel are collectively called aerosol. Some of the particles are originating from nature. Different pollutants have an effect on air which takes part to decrease the air quality [47]. Figure 1 represents the ratio of various pollutants examined in earlier proposed approaches. The commonly used pollutants are PM2.5, PM10,SO2, NO2, O3, CO. In this chart, we analyze approximately 40 researches that use different pollutants in air pollution prediction approaches.

Air Quality Index (AQI) is an index that gives a stage of air pollution-related with its health effects. The AQI focuses on several fitness outcomes that humans may experience based on the level and hours of publicity to the pollutant concentration. The AQI values are exclusive from country to country on the basis of the air high-quality standard of the country. The higher the AQI level higher is the danger of health-related problems [28]. Figure 2 shows the air quality index levels.

To check whether the air is safe or unhealthy and to understand the quality of air, AQI is used. It is, therefore, necessary to take preventative measures to forecast AP parameters, like ozone and nitrogen dioxide [52].Those parameters also have a direct influence on the degree of AP and are considered by sensing equipment installed in many SCs. The incorporation of devices into cities produce a large quantity of timely annotated data. The Deep Learning (DL) principle is highlighted by rectifying and forecasting a vast volume of data.

The main reasons for Deep Learning's popularity are increased chip processing capabilities, significantly reduced computer hardware costs, and the latest progress in ML and

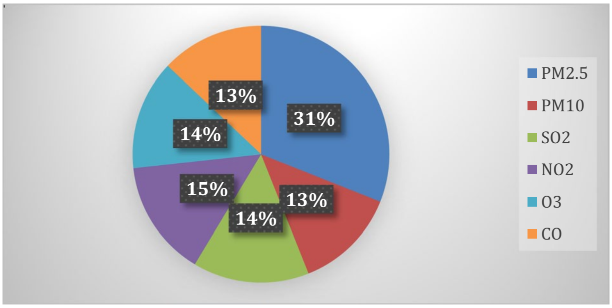


Fig. 1 The ratio of different pollutants was examined in earlier approaches

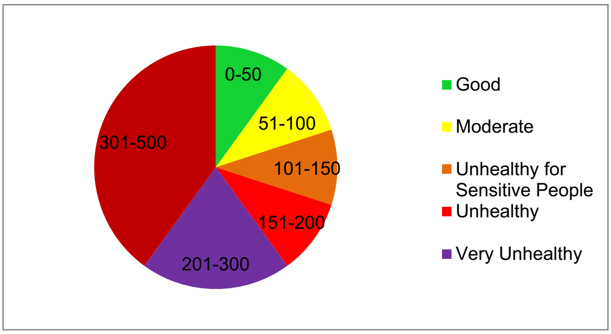


Fig. 2 AQI Levels

information processing studies [11]. DL is therefore of great interest in educational and industrialized fields and has been successfully useful in various fields, like in classification, object detection, dimensionality reduction, etc. Monitored gases, therefore, generate sequential data which can be analysed at an early stage to guess future value [35].

Figure 3 is a simple model compared to more complex models that require supercomputer processing power to generate global air quality forecasts. However, it has the advantage of visually illustrating the fundamental concept of air quality forecasting.

Forecasting is the practice of making accurate predictions about unknown conditions based on historical evidence. One of the most scientifically and technologically hard topics in the last century has been real-time measurements and weather forecasting. Weather forecasts are frequently generated by gathering quantitative data on the present condition of the atmosphere and applying a scientific understanding of atmospheric processes to estimate how the atmosphere will change in the future, which can then be saved as observation data. Feedback is a method by which a computer system interprets data about the performance and uses it to improve the quality of their work or learning tactics. It's the process of analyzing your past performance and taking steps to improve your future performance.

This information is used to forecast future patterns by looking at data across several years and discovering patterns. The system might be able to convert the patterns in the graph into a formula that can reliably forecast what will happen in the future. Pattern forecasting is a type of quantitative forecasting that is based on measurable, real data from the past. It makes use of time series data, for which the numerical value is identified at multiple points in time and is sent to the forecast engine. Inventory management gains a new dimension due to the forecasting engine. It creates forecasts using tried-and-true statistical forecasting methods, taking into account a variety of important factors such as historical data, seasonal changes, and promotional efforts. It makes air quality forecast data available to the public after improving prediction accuracy. When more precise data is required, it is saved as project data, which can be used to learn by comparing observed data and identifying recurring patterns.

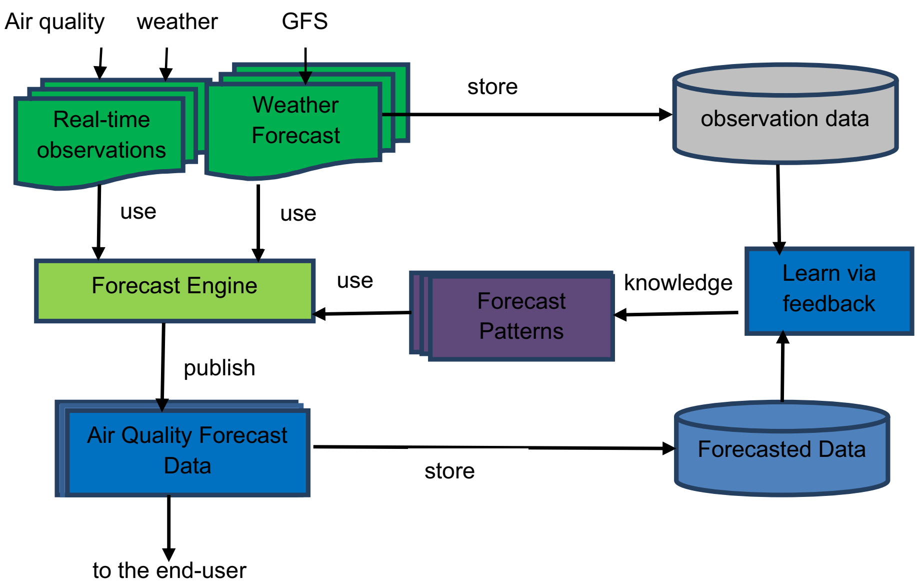


Fig. 3 Air quality forecasting system

Finally, air quality forecasting systems have been the subject of numerous research studies. The objective is to "learn" by comparing actual and anticipated data and looking for trends that repeat.

A compatible combination of two components is required for an AQ forecasting system, just as it is for a weather forecasting system. These components include a set of predictive methods adapted to the demands of the client community, as well as an observation network proficient in supplying real-time atmospheric composition measurements required to initialize the models and assess the forecast's quality [37].

The interrelationship between the primary elements of an air quality forecasting system is depicted in Fig. 3. A well-made forecast system contains a method for producing the forecastations as well as an observation system for evaluating the forecast's quality and identifying areas for improvement. Any prediction system that is created AQ prediction must have a comprehensive observing system. It can only give confidence to future estimates and finds locations where enhancements are desired by comparing actual and anticipated pollution levels.

1.1 Challenges in air quality monitoring and forecasting

While progressive studies are aimed at the successful monitoring of air quality in real-time, some challenges still need to be addressed [18]:

1. Implementation of more reliable, practical real-time AQ monitoring technologies that will provide precise concentrations by taking into account various meteorological conditions.
2. The system's instability and nonlinearity must be explored.
3. Combining the devices into an internet infrastructure that is manageable at all times and from anywhere.
4. The need to improve short and long-term forecasts by using continually monitored air quality data and taking into account all elements that influence them.
5. Hybridization or combination to produce the finest performance.

In air quality forecasting techniques, statistical DL approaches are used, which use a huge quantity of historical data with the assumption that the elements are connected. Statistical methods focus on statistical techniques rather than an elaborate theoretical methodology to predict air quality. Figure 4 depicts the general Air Quality Forecasting Model. It uses an empirical method to anticipate future air quality indexes based on a threshold value, a past AQI, or some weather conditions. This method, on the other hand, has the issue of being unable to handle abrupt changes in air quality induced by an unknown environmental change.

The purpose of the AQ monitoring station is to track the concentrations of specific air pollutants at different times of the day. The device collects data on AP and transmits it to a remote server for logging and analysis. Various gas sensors are included in the gadget. Apart from gas sensors, the gadget also incorporates dust particle sensors for detecting suspended particulate materials of PM10 and PM2.5 sizes in the air. Temperature, relative humidity, barometric pressure is also included in the device for reliable pollution parameter measurement and analysis. AQ data is sent to the cloud for analytics via cellular connectivity, and the data is displayed in the form of an AQI in a web application.

Analyzer's for pollution measurement are included in the continuous online ambient air quality monitoring systems. A Comprehensive Operation and Maintenance Contract covers the Air Quality Monitoring System (COMC). EMRC, IMD checks the operation of all

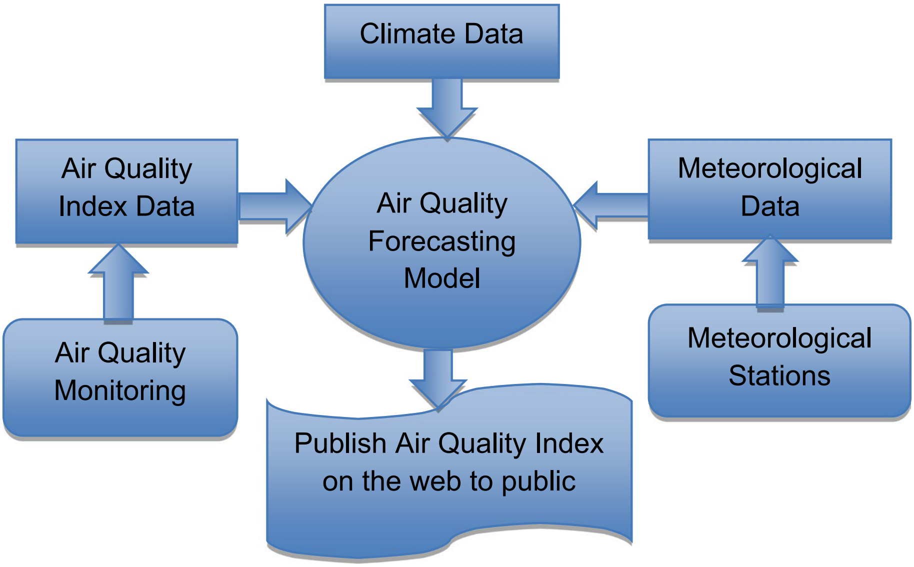


Fig. 4 Abstract view of air quality forecasting model

AQMS daily. The control room receives AQ and meteorological data collected from monitoring stations. The AQMS data is converted to AQI by the Central Control Room, which then sends the AQI together with meteorological data and Air Quality forecasts to each city's FTP site. The data is subsequently sent from the FTP server to the DDS system's control server, which finally sends it to the Digital Display Boards in each city [30].

1.2 Deep learning concept

Deep networks are effective in learning knowledge representations in high-dimensional space with the context of weights connections among neurons in the network. Thus, network architecture, or network organization, is scheming of the knowledge space of different domains. DL is used successfully in various fields, like computer vision [4], natural language processing [38], speech recognition [21], and then in basic science like physics and chemistry [34]. The same level of achievement may be predicted for air quality (AQ)forecasting, as the growth of contaminants can be compared to the prediction of image sequences [9].

Above the previous centuries, the quantity of air quality information produced has increased exponentially. For example, new global images are continuously generated by satellite remote sensors, while improved computational power allows us to create routine simulations with mathematical models. The capability to gather and construct facts, however, far outweighs our capability to interpret them. Based on deep learning frameworks we may model AQ systems to utilize such large data for forecasting [33]. Since these systems are significantly complex, spatially expensive, and evolutionarily heterogeneous. DL can detect high-dimensional data automatically, process data with dimensionality reduction, and extract low-dimensional data [13].

1.3 Related work

For air quality monitoring several approaches over the years have been developed. Jianxian Cai et al. [5] formed a noise reduction Auto Encoder (AE) through LSTM to extort the inherent AQ features of originally monitored data. The structure of LSTM in the DAEDN approach was sketched as Bi-LSTM for solving the issue of a lag in the unidirectional LSTM, estimated outcomes and so to additionally expand the accurateness of the model. The suggested approach for executing AQ prediction in comparison to the (STANN), (ARMA) [5], and (SVR) [45] models gives superior performance [31]. Junxiang Fan et al. [17] developed a new DL framework that can effectively handle missing values in ST forecasting tasks and presented a DRNN [12] set up through LSTM [19] over the top of fixing approaches. The suggested approach can forecast both unexpected massive pollution and average patterns through comparatively high precision. This study can predict AQ for up to 48 h by combining multiple NN with ANN [48], CNN [40], & LSTM to take out ST relations. Ping-Wei Soh et al. [46] proposed a prognostic model with several meteorology data for some hours in addition to related information of the elevation space to extract terrain impact on AQ. Brian S. Freeman [14] presented an approach in which 8 h averageO3 concentrations were predicted using DL comprising RNN with LSTM. An LSTM deep NN& a (SARIMAX) [50] were directed on meteorological time series &AQ data. The implementation of LSTM in the prediction of PM2.5 for 24 h was estimated and compares after considering the best configuration of both algorithms. Yanlin Qi et al. [39] approached a DL hybrid model which integrates Graph CNN and LSTM (GC-LSTM) model to forecast PM2.5 ST variation. S. Jeya&Dr. L. Sankari [23] attempt to forecast one of the harmful diseases trigger around the world as PM2.5 pollutant by using bidirectional LSTM model. In comparison to the existing model, the proposed approach is superior by estimating RMSE, MAE, and SMAPE. For analyzing the PM2.5 correlation and other supplementary data, the Pearson Correlation Coefficient is employed. The extracted features were useful to construct a deep ensemble network (EN) [15] that integrates the RNN, LSTM, and GRU network [26] for predicting PM2.5. The grouping of CNN and LSTM automatically operates both the spatial and temporal features data. Van-Duc Le et al. [29] introduce the way to convert the polluted data into an image sequence that leverages the utilization of the Conv LSTM approach to exclaim and forecast AQ for the whole city at an identical time. ThanongsakXayasouk et al. [53] predict fine PM concentrations using LSTM and deep AE approaches, compares the outcomes in provisions of RMSE. They applied the model for hourly basis data from 25 stations for the time duration 1 January 2015 to 31 December 2018. Kaya&Şule [24] have hourly real-world data from Istanbul for air pollution prediction. The suggested DFS model is hybrid & flexible comprising LSTM and CNN. This is proficient in generalization with standard and versatile dropout layers. Bo Zhang [58] proposed a DL model based on an AE and Bi-LSTM for PM2.5 concentration to disclose the correlation between PM2.5 and other climate variables. Luo Zhang [57] proposed a semi-supervised approach for PM2.5 concentration forecasting. The method uses empirical mode decomposition (EMD) and BiLSTM neural networks. For estimating and monitoring PM2.5, CNN and LSTM are integrated and apply to the PM2.5 forecasting system [20]. Table 1 shows the earlier studies conducted for air pollution prediction with deep learning.

1.4 Spatiotemporal (ST) based forecasting models

Dewen Seng et al. [42] suggested a forecasting model based on LSTM with MMSL of supervised learning in which the particle concentration data were integrated for the same time. Joseph & Agustí [41], on the basis of a dynamic linear modelling structure with Bayesian inference, proposed an ST model which was functional in CO, SO2, O3, NO2, and PM 2.5. The approach was completed and enhanced the traditional approaches in air pollution modeling. Xiaolu Zhou et al. [60] offered a novel ST interpolation model by combining data fusion techniques and LSTM RNN, capable to attain high estimation accuracies and predicted the PM 2.5 values on ST features for monitored data. Xiaotong Sun et al. [49], for PM 2.5 prediction on an hourly basis proposed an ST-GRU, an extension of RNN including ground pollution monitoring GPM, FE, SMM variables. For analysing the time series dependence of PM 2.5 they used the RNN characteristics. Lianfa Li [32] developed AE-based residual DN to make strong accusations of (MAIAC-AOD) retrieves satellite aerosol optical depth was done and evaluation of PM2.5 at a high spatiotemporal resolution. For capturing the complex ST relation of meteorological data, Lie Zhang et al. [59] presented a deep CNN (ST-OR-ResNet) for air forecasting. The method was used to improve prediction accuracy while also increasing training performance.

Graph 1 and 2 shows the RMSE and MAE respectively for various earlier proposed approaches worked on an hourly basis. From these graphs, we can observe that the LSTM model has the lowest and the CNN-LSTM model have the highest RMSE and MAE values.

1.5 Deep learning terminologies

1.5.1 Recurrent Neural Network (RNN)

It's a form of network in which the precedent phase output is used as an input in the present phase. In a conventional neural network, the I/O are self-sufficient of one another, but in certain situations, such as when predicting the subsequent word of a sentence, the earlier words are necessary, and therefore the previous words must be remembered. RNN was created with the aid of a hidden layer to solve this problem. The key and very significant feature of RNN is the hidden state that recalls little information about a chain. It has a "memory" that retains all of the measurements' data. This utilizes similar parameters for every input because it plays the identical function on all inputs or hidden layers to generate the output. Unlike other networks, this one lessens parameter complexity [1].

RNNs can process sequential data and provide sequential output. RNN employs a nonlinear activation function at each unit to learn the input-output relationship. The power of an activation function is greater than that of a linear activation function. It updates and adjusts network weights using the Backpropagation algorithm during the learning phase [41]. From the network output, the BPTT works backward through the network layer by layer, changing the weights of each layer based on the layer's computed percentage of the overall output error. The shifting of weights denotes the distance between the current output and the anticipated outcome. During the weights update, two major difficulties can occur exploding gradient and vanishing gradient.

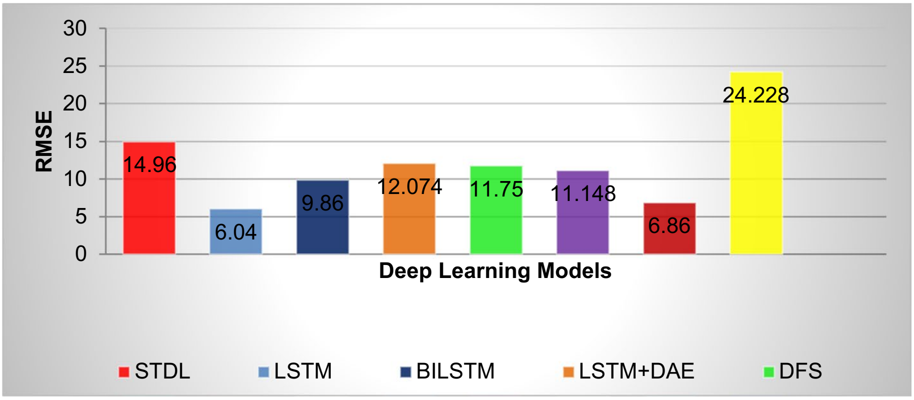
RNNs introduce FNN-based cyclic self-connection of neurons into the network. The entered data can thus be recognized, and data sequences can affect network outputs by neurons that are self-connected. Getting the benefit of their memory properties, in many applications, RNNs outperform FNNs. RNNs, however, may not succeed to catch long-term

Table 1 The earlier studies conducted for air pollution prediction with deep learning

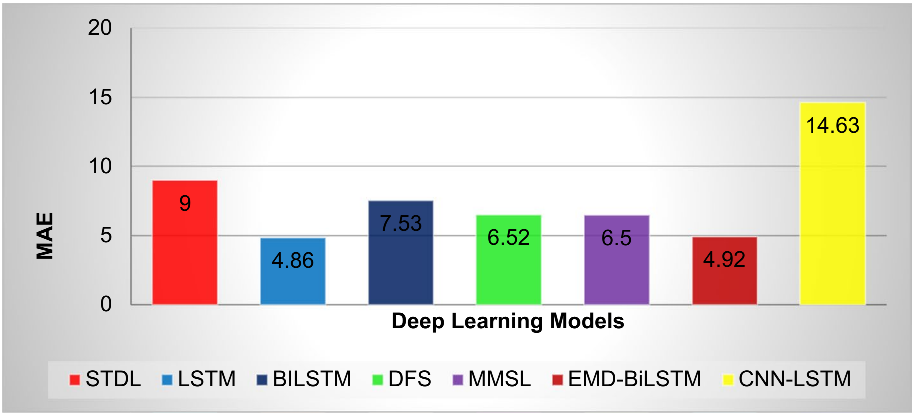
Author	Method Used	Pollutants	Time Granularity	Prediction Performance
Xiang Li [31]	STDL Model	PM2.5	Hourly basis	RMSE = 14.96, MAE= 9.00, MAPE= 21.75 %
Junxiang Fan et al. (2017)	DRNN model	PM2.5	8 h	IA= 0.936, MAE= 23.721, RMSE= 35.7362
Ping-Wei Soh et al. [46]	ST-DNN model	PM2.5	48 h	MAE = 7.668
Brian S. Freeman et al. [14]	(RNN+LSTM) model	O3	8 h	MAE= 0.41
Thawephol & Nuwee (2019)	LSTM model	PM 2.5	Hourly basis	RMSE= 6.04 MAE= 4.86
Yanlin Qi. et al. [39]	GC-LSTM model	PM 2.5	72 h	IA= 0.92, MAE= 24.21, RMSE= 38.83
S.Jeya, Dr. L. Sankari [23]	BILSTM model	PM2.5	Hourly Basis	RMSE = 9.86, MAE = 7.53, and MAPE = 0.1664.
Jianxian Cai et al. [5]	DAEDN model	PM2.5, PM10, SO2, NO2, O3, CO	24 h	RMSE= 15.504 MAE= 6.789
Canyang Guo et al. [15]	(RNN+LSTM+GRU) model	PM2.5	Hourly Basis	MAE= 6.19 MAPE= 16.20 %
Van-Duc Le et al. [29]	ConvLSTM model	PM2.5	12 h	RMSE= 8.59883
Thanongsak Xayasouk et al. [53]	(LSTM+DAE) model	PM2.5, PM10	Hourly basis	For LSTM, RMSE= 12.074 For DAE, RMSE= 16.528
Kıymet Kaya & Gündüz Öğüdücü [24]	DFS model	PM 10	Hourly basis	MAE= 6.52 RMSE= 11.75
Dewen Seng et al. (2020)	MMSL model	PM2.5, CO, NO2, O3, SO2	Hourly basis	MAE= 6.50 RMSE= 11.148
Joseph, Agust Foguet [41]	Dynamic Linear Model	CO, SO2, O3, NO2, PM2.5	Hourly	RSE= 1.1086 NSE= -0.111
Bo Zhang et al. (2019)	AE-Bi-LSTM model	PM 2.5	24 h	RMSE= 3.294
Luo Zhang et al. (2021)	EMD-BiLSTM model	PM 2.5	Hourly	RMSE= 6.86 MAE= 4.92
Chiou-Jye Huang et al. (2018)	CNN-LSTM model	PM 2.5	Hourly	RMSE= 24.228 MAE= 14.63

dependency in entered data, and when the training time is too long, they may look at the issue of disappearing and exploding gradients [33]. Figure 5 [6] shows the structure of RNN. To combat vanishing and exploding gradients, several techniques have been offered in the literature. LSTM and GRU are the most common, as stated next.

Steps to be performed in RNN-

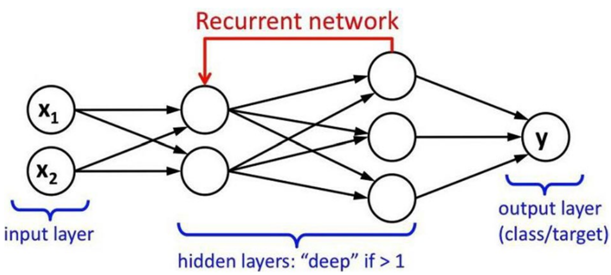


Graph 1 RMSE (Root Mean Square Error) values for various earlier proposed approaches worked on an hourly basis



Graph 2 MAE (Mean Absolute Error) values for various earlier proposed approaches worked on an hourly basis

Fig. 5 Recurrent Neural Network (RNN)



1. The first input x_t is given as input.
2. Compute the current state h_t and then h_t becomes h_{t-1} for the next time step.

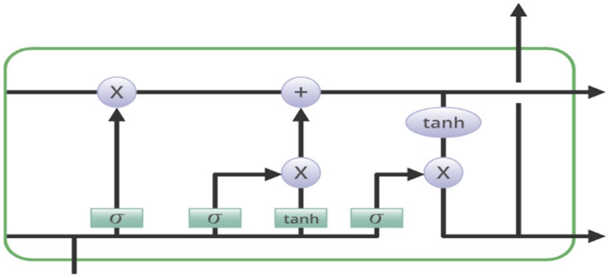


Fig. 6 Long- Short Term Memory (LSTM)

$$h_t = f(h_{t-1}, x_t) \tag{1.1}$$

3. Depending on the situation, one can continue through as many time steps as necessary, combining the data from all previous steps.

$$h_t = tanh\left(W_{hh}h_{t-1} + W_{xh}X_t\right) \tag{1.2}$$

4. When all steps were completed calculate the output y_t with the final current state and then compared to the actual output and the error is generated.

$$y_t = W_{hy}h_t \tag{1.3}$$

5. The error is then backpropagated to the network, which updates the weights and trains the network.

1.5.2 Long- Short Term Memory (LSTM)

LSTM is a type of RNN. This is primarily based on the gradient descent approach and can cut the gradient where this does not harm. The multiplicative gate units, work like a valve that can figure out how many records need to omit and how many of them must be discarded. In RNN, the result from the last step is fed as input in the contemporary step. This handles the issue of long-term dependencies of RNN whereas this can't predict the data saved in the long-term memory, although can supply greater precise prediction from the latest data. When the gap will increases, RNN no longer gives proficient results. By using the default, LSTM maintains the records for a lengthy phase of time. It is used for process, prediction, and classification based on time sequence data [8].

The memory block of LSTM contains three kinds of nonlinear multiplicative gates: the input, output, and forget gate. Figure 6 shows the structure of LSTM. The gates controls the memory operations and decide whatever or not the enter records have to be recognized. The input gate maintains the flow of cell activation enters into a memory cell, whereas the output gate maintains the flow from a memory cell to other nodes.

The process is split into two parts. The first part, where the values are to be updated, is an Input Gate Layer. A hyperbolic tangent (tanh) layer will produce Ct in the second part, a list of new candidates values. As seen in Eq. (2.3). The last step is to integrate the previous two components to generate a state update. The input gate serves to determine how much information can be added to the cell state from the current input. The old cell state Ct-1 is updated into the new cell state Ct. As represented in Eq. (2.4). Instead of the forget process, the old state is multiplied by ft, so that the network can forget information that is no longer needed as shown in Eq. (2.1). The final step is to find out what the network output will be. The result is derived from the cell state, but with several changes. The cell state parts are determined using a sigmoid layer. Those values must be in the range [- 1, 1] and a hyperbolic tangent is applied to the cell state. These values are then multiplied by the sigmoid gate output so that the final output is represented purely by one part of it as in Eqs. (2.5) and (2.6).

The gate units receive a value between 0 and 1. The value 0 means that no facts should pass while the value 1 suggests that the complete flow of records has to pass. All the values in the middle between 0 and 1 point out what is the volume of information that passes. During the learning process, they learn to close and open to adjusting the constant error flow of the cell state.The advantage of LSTM networks is to instruct lengthy time series and performs higher than conventional RNN in several applications [33]. The computation complexity of LSTM in terms of space and time is O (1) [36].

Forget Gate- This decides which data we are going to set from the cell state.

$$f_t = \sigma(W_f * \left[h_{t-1}, x_t\right] + b_f) \tag{2.1}$$

Input Gate-This decides what latest data we are going to store in the cell state.

$$i_t = \sigma(W_i * \left[h_{t-1}, x_t\right] + b_i) \tag{2.2}$$

$$C_t = tanh(W_C * \left[h_{t-1}, x_t\right] + b_C \tag{2.3}$$

Update- Update, scales by what we decide to update.

$$C_t = f_t * C_{t-1} + i_t * C_t \tag{2.4}$$

Output Gate - Output based on the updated state.

$$O_t = \sigma(W_o\left[h_{t-1}, x_t\right] + b_o \tag{2.5}$$

$$h_t = O_t * tanh\left(C_t\right) \tag{2.6}$$

1.5.3 Gated Recurrent Unit (GRU)

A simplified variant of the LSTM networks are the GRU network. It comprises only three gates, unlike LSTM, it does not hold an internal cell state [16]. The information enclosed in an LSTM recurrent unit in the internal cell state is inserted into the GRU secret state. This collective data is agreed on to the subsequent Recurrent Gated Unit. The gates and weights of the two approaches are the most significant variations between them. GRU has a faster learning process, fewer gates, and a shorter memory than LSTM. The different gates of a GRU are stated here-.

Update Gate: It decides what information needs to be imported into the future from the past. In an LSTM recurrent unit, this is similar to the output gate. Equation (3.1) represents how the update gate controls new data and information generated by earlier activations.

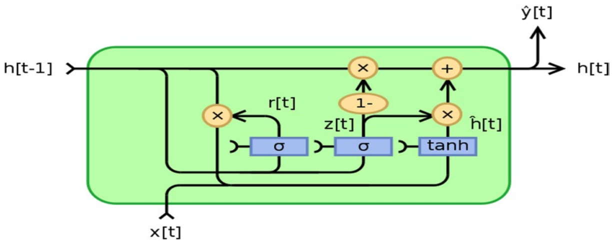


Fig. 7 Gated Recurrent Unit (GRU)

Reset Gate: It decides what to forget from experience. In an LSTM recurrent unit, it is equivalent to the combination of the input gate and the forget gate. Equation (3.2) represents which reset gate activity is involved in candidate activation.
Current Memory Gate: The Input Modulation Gate is a component of the Input Gate, it is integrated into the Reset Gate and is utilized to establish several non-linearity into the input and also to render the input zero-mean. Another purpose for making it a component of the reset gate is to decrease the influence on the existing knowledge that is being transferred into the future that past data has on it [26]. Equation (3.3) and (3.4) obtains the output of the current state.

The formula for calculating the next output and state value in the GRU is-

$$z_t = \sigma\left(W_z * [x(t), h(t-1)]\right) \tag{3.1}$$

$$r_t = \sigma\left(W_r * [x(t), h(t-1)]\right) \tag{3.2}$$

$$h(t) = \sigma\left(W_h * \left[x(t), \left(r_t * h\right)(t-1)\right)]\right) \tag{3.3}$$

$$h(t) = \left(1 - z_t\right) * h(t-1) + z_t * h(t) \tag{3.4}$$

Where,
σ = activation function.
x(t) = input.
h(t-1) = previous output.
W_z, W_r, W_h= weights
Compared to LSTM, the benefit of using GRU is that it has fewer parameters and therefore less training machine loads. Nevertheless, performance on music and speech signals has been shown by the GRU networks as LSTM or improved performance on smaller datasets. Figure 7 [7] shows the structure of GRU.

1.5.4 4. Deep Belief Networks (DBN)

DBNs are probabilistic generative models that are built up from multiple layers of Restricted Boltzmann Machines (RBMs), each of which has a visible layer and a hidden layer. Each layer of the RBM is trained until convergence, and then the product of the machine's output layer is then fed as an input in the series to the next Boltzmann machine,

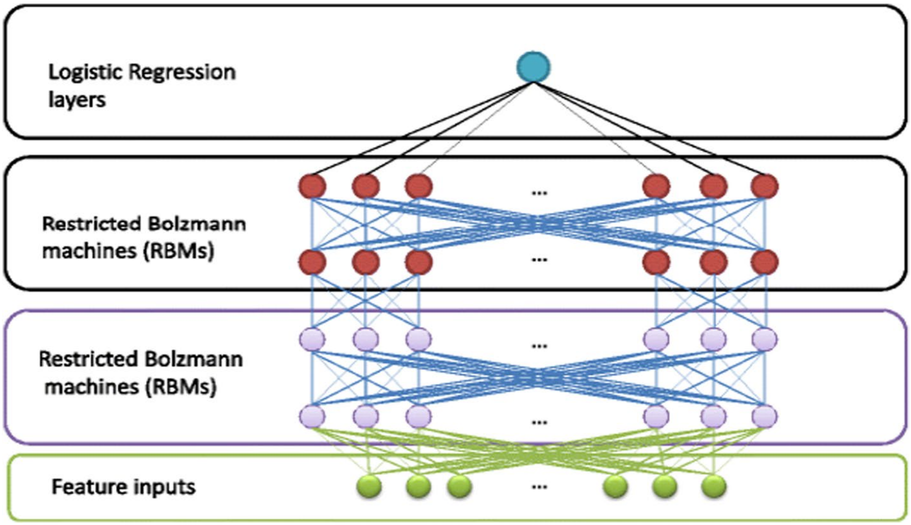


Fig. 8 Deep Belief Networks (DBN)

which is then, trained itself until convergence, and so on until the entire system has been trained. The full distribution of DBN is shown in Eq. (4.1).

To initialize the network parameters of each node, the Restricted BM conducts unsupervised pre-training layer after layer from bottom to top. A SoftMax classifier is set after the pre-training phase in the final layer of DBN for feature classification [3]. Ultimately, the whole network is optimized and tracked using the backpropagation algorithm tuned through the labeled network. Also, a tenfold cross-validation approach is widely used to measure the efficiency of the model and assess the overfitting of the model. Figure 8 [44] shows the structure of DBN.

The full distribution of DBN is as follows-

$$p\left(x, h^{(1)}, h^{(2)}, h^{(3)}\right) = p\left(h^{(2)}, h^{(3)}\right)p\left(h^{(2)}\right)p\left(h^{(1)}\right) \tag{4.1}$$

Where,

$$\mathrm{p}\left(\mathrm{h}^{(2)}, \mathrm{h}^{(3)}\right) = \exp\ \exp\left(\mathrm{h}^{(2)\mathrm{T}}\mathrm{W}^{(3)} + \mathrm{b}^{(2)\mathrm{T}}\mathrm{h}^{(2)} + \mathrm{b}^{(3)\mathrm{T}}\mathrm{h}^{(3)}\right)/\mathrm{Z}$$
$$\mathrm{p}\left(\mathrm{h}^{(1)}\middle|\mathrm{h}^{(2)} = \sqcap_{\mathrm{j}}\mathrm{p}\left(\mathrm{h}_{\mathrm{j}}^{(1)}\middle|\mathrm{h}^{(2)}\right)\right.$$
$$\mathrm{p}\left(\mathrm{h}^{(1)}\right) = \sqcap_{\mathrm{i}}\mathrm{p}\left(\mathrm{x}_{\mathrm{i}}\middle|\mathrm{h}^{(1)}\right)$$

1.5.5 Auto Encoder (AE)

An Auto Encoder is a form of NN that encodes input data for output data reconstruction [55]. To begin, the AE first acquires to take the input's important features. Figure 9 depicts an AE with different layers. To train set x(1).x(n) so that x(i)€Rd, the autoencoder model's initial step is to encode the single input to the hidden layer, as shown in Eq. (5.1); then the output is shown in Eq. (5.2).

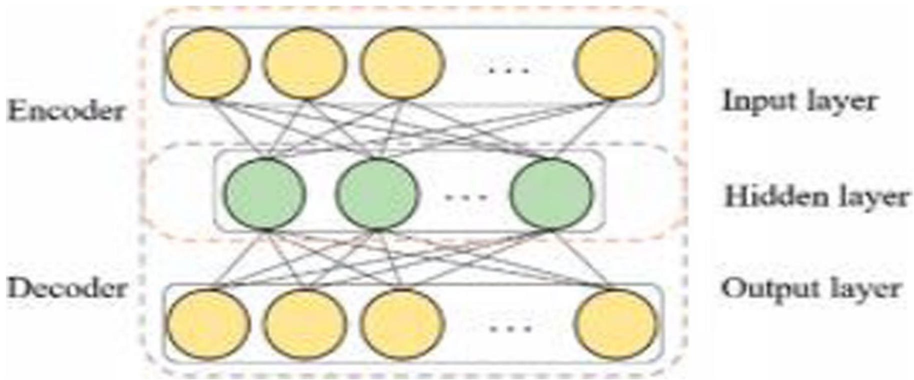


Fig. 9 Processes of the autoencoder model

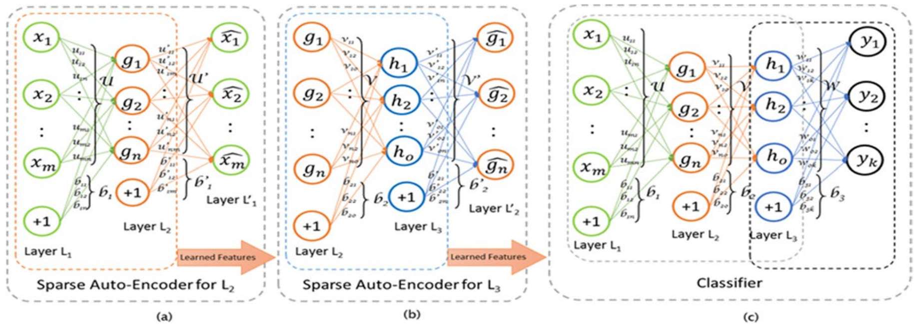


Fig. 10 Stack Auto Encoder(SAE)

$$y(x) = f(W_1 x + b) \tag{5.1}$$

$$z(x) = g(W_2 x + c) \tag{5.2}$$

where W_1 is the optimization process's weight matrix, b is the encode bias vector, W_2 is the output layer's decoding matrix, and c is the decoding bias vector.

An AE uses backpropagation to produce an output value that is almost identical to the input value and is an unsupervised learning algorithm. An SAE is a network consisting of many sparse layers. AE connecting the result of every hidden layer to the input of the hidden layer in succession. Figure 10 [55] shows the structure of Stacked AE comprises three stages [27]. By an unsupervised algorithm, the hidden layers are trained and after that fine-tune by a supervised process. The Stacked AE uses input data to instruct the autoencoder and obtain the learned data.

- For the next layer, the trained data from the earlier layer is used as an input and this continues until the training is done.
- With the use of the backpropagation algorithm, to lessen the cost function until all the hidden layers were trained and weights are changed with the training set to accomplish fine-tuning.

1.6 Analysis of approaches used in air pollution prediction

Leading to the quick growth of industrial technologies, the global environment is becoming more contaminated. However, the latest air quality forecasting methods of ML are unsuccessful to examine the reasons for the shift in the concentration of AP since many of the forecasting approaches are more based on model selection. The study of time series data has been considered for decades by analysts. There is no potential to model sequential data with elevated accurateness in typical shallow models. DL is also correlated with complex real-world data, as time-series data is complicated, high-dimensional, and noisy and have been useful in time series prediction problem. As the structure of the latest DL is much flexible, to fit the dataset, the model could be deep and complex. In a single deep NN model, over-fitting problems may exist whenever the no. of weights in the network is huge.

NN models are a single sequence element for every neuron of the input layers of networks, whereas each neuron of the input layer in RNN, LSTM, and GRU is a vector encoded by elements of the earlier sequence. Further DL approaches are useful in prediction issues, e.g. RNN and LSTM models, in addition to NN models. A variant of the FNNis the RNN. These models sequence data and take into account the 'memory' transfer, so its training process has to stack the results of BP over the time dimension which results in the BPTT algorithm. The disadvantage of using RNN is the gradient may tend to be 0 through the extension of time steps, leaving it difficult to train the parameters of a long-term dependent network. The problem is called the 'vanishing gradient'. For learning long-term dependence, LSTM, an extended RNN model, differs from RNN because there is a gradient disappearance phenomenon in RNN. The gates make sure that LSTM gradient information will not disappear through backpropagation, enabling LSTM to learn long-term dependencies. LSTM parameters are trained by using BPTT. In internal structure, GRU and an extended LSTM model differ from LSTM because it has three gates and GRU has two.The performance of the GRU model for predicting PM10 concentration is found to be somewhat superior to the RNN and LSTM models. The structure of the LSTM network was planned as Bi-LSTM to overcome the issue of a lag in the outcomes of the unidirectional LSTM prediction and thus additional enhance the model's prediction accuracy. Bidirectional LSTM, however, retains both past and future information at a single point in time with the aid of two combined hidden states.

For the spatiotemporal prediction of air quality, many researchers try to use STDL models. One of the temporal and spatial associations can be considered by the classic models. The analysts have therefore planned several new joint models for AQ forecasting to enhance the accuracy of the model prediction. The model uses CNN to remove the spatial features of inputs among monitoring stations, a new combined forecasting system based on CNN and LSTM for PM2.5 concentration and uses LSTM to forecast future concentrations of AP by learning the features found in conventional time-series data of air pollution concentration. Autoencoder is a specific type of NN with a symmetrical structure, in an unsupervised or supervised way, from encoding to decoding layers and the same input and output variables. It is intended by the middle latent layer to learn effective data compression or latent coding representation, which can enhance learning performance.

Deep Learning techniques can be effective in extracting representative air quality features for air pollution forecasts without prior information, resulting in improved prediction performance. RNNs are commonly utilized in air pollution forecasting studies [2, 43]. It works well with any sort of sequential data analysis. It was created to model time series pollution datasets sequentially. Because data in air pollution forecasting evolves, and RNN can learn changes in the time domain, it may be a better prediction solution. Though it can handle large datasets, it suffers from gradient descent difficulties, necessitating the development of a new variant of the RNN structure.

The LSTM algorithm is commonly used to anticipate air pollution. Its primary job is to recall information for extended periods, with internal memory for processing sequences of inputs, and to record old and current air pollution data. The model uses a framework to anticipate air quality, as well as pollution and meteorological variables from time-series data. To solve long time lag difficulties in air quality forecasting models, LSTM can control noise, spread patterns, and constant variables, as used in the study [10].

In terms of computing time and performance in forecasting algorithms, GRU outperforms LSTM [51]. The spatial-temporal method is used by GRU models to create PM 2.5 prediction models. GRU is used to train a prediction model using data from each season. It utilizes less memory, runs faster, and trains faster than LSTMs in forecasting approaches since it has fewer training parameters. The GRU model was used to forecast the concentration of air pollutants based on the features of high-dimensional and complicated air pollution time series.

DBN is used to extract stronger feature representations, and shared representations are utilized to tackle numerous prediction-related tasks at once. A model that can be pretrained by a deep belief network for forecasting nonlinear systems is presented and tested on the forecast of air quality time series using DBN's powerful representational ability. A DBN is used to solve classification and prediction problems, made up of numerous RBM layers and one back-propagation layer [25].

In recent air pollution forecasting studies, autoencoders have been employed as pre-training techniques to extract representative spatiotemporal properties. An SAE model is utilized in many studies to derive realistic spatiotemporal air quality parameters. The stacked autoencoder might give essential auxiliary information for PM 2.5 time series prediction by encoding the key evolution pattern of urban meteorological systems [54].

2 Conclusions

External air pollution is now severely affecting the health of human life in major cities. For observing air pollutants, several countries have installed monitoring stations within major cities. Air pollution sources may come from manufacturing, people, cars, or natural sources. As a result, several complex factors influence air pollution, and forecasting air pollution is a difficult problem.

Air Quality monitoring and forecasting has recently become a popular topic of research, with rapid progress being made. Using a range of statistical methods and techniques, several countries have created their monitoring and forecasting models. The latest developments in air quality forecasting and real-time monitoring were discussed in this paper. Different forms of large data, like images, sounds, and numeric data were used to forecast air pollution with deep learning. Some examples of methods used for this purpose are LSTM, STDL, GRU, and CNN. Their efficacy is primarily determined by the data and algorithms they employ. When deep learning models for air pollution evaluation are compared to other approaches such as ANN and fuzzy logic, deep learning models come out on top.

References

1. Introduction to recurrent neural network: https://www.geeksforgeeks.org/introduction-to-recurrent-neural-network/. Accessed Mar 2021

2. Athira V, Geetha P, Vinayakumar R, Soman KP (2018) DeepAirNet: applying recurrent networks for air quality prediction. Procedia Comput Sci 132:1394–1403. https://doi.org/10.1016/j.procs.2018.05.068
3. Notes on deep belief networks: https://www.kaggle.com/residentmario/notes-on-deep-belief-networks. Accessed Mar 2021
4. A Gentle introduction to computer vision. Machine learning mastery: https://machinelearningmastery.com/what-is-computer-vision/. Accessed Mar 2021
5. Cai J, Dai X, Hong L, Gao Z, Qiu Z (2020) An air quality prediction model based on a noise reduction self-coding deep network. Math Probl Eng 2020. https://doi.org/10.1155/2020/3507197
6. Chaterjee CC (2019) Implementation of RNN, LSTM, and GRU. Towards data science. https://towardsdatascience.com/implementation-of-rnn-lstm-and-gru-a4250bf6c090. Accessed Mar 2021
7. Chao K et al (2014) Learning phrase representations using RNN encoder-decoder for statistical machine translation. https://arxiv.org/abs/1406.1078
8. Chugh A (2019) Deep learning | introduction to long short term memory. https://www.geeksforgeeks.org/deep-learning-introduction-to-long-short-term-memory/. Accessed Mar 2021
9. Cigizoglu HK, Alp K, Kömürcü M (2005) Estimation of air pollution parameters using artificial neural networks. Adv Air Pollut Model Environ Secur (1):63–75. https://doi.org/10.1007/1-4020-3351-6_7
10. Delgado A, Acuña RMR, Carbajal C (2020) Air quality prediction (PM2.5 and PM10) at the upper hunter town - Muswellbrook using the long-short-term memory method. Int J Adv Comput Sci Appl 11(4):318–322. https://doi.org/10.14569/IJACSA.2020.0110443
11. Deng L (2014) A tutorial survey of architectures, algorithms, and applications for deep learning. APSIPA Trans Signal Inf Process, vol 3, no 2014. https://doi.org/10.1017/ATSIP.2013.99
12. DRNN. Deep recurrent neural network. Dive into deep learning. https://d2l.ai/chapter_recurrent-modern/deep-rnn.html. Accessed Mar 2021
13. Donnelly A, Misstear B, Broderick B (2015) Real time air quality forecasting using integrated parametric and non-parametric regression techniques. Atmos Environ 103:53–65. https://doi.org/10.1016/j.atmosenv.2014.12.011
14. Freeman BS, Taylor G, Gharabaghi B, Thé J (2018) Forecasting air quality time series using deep learning. J Air Waste Manag Assoc 68(8):866–886. https://doi.org/10.1080/10962247.2018.1459956
15. Guo C, Liu G, Chen CH (2020) Air pollution concentration forecast method based on the deep ensemble neural network Wirel Commun Mob Comput 2020. https://doi.org/10.1155/2020/8854649
16. Gupta A (2019) Gated recurrent unit networks. https://www.geeksforgeeks.org/gated-recurrent-unit-networks/. Accessed Mar 2021
17. Fan J, Li Q, Hou J, Feng X, Karimian H, Lin S (2013) A spatiotemporal prediction framework for air pollution based on deep RNN. ISPRS Ann Photogramm Remote Sens Spat Inf Sci 4:15–22. https://doi.org/10.5194/isprs-annals-IV-4-W2-15-2017
18. Hable-Khandekar V, Srinath P (2017) Machine learning techniques for air quality forecasting and study on real-time air quality monitoring. 2017 Int. Conf. Comput. Commun. Control Autom. ICCUBEA pp 1–6. https://doi.org/10.1109/ICCUBEA.2017.8463746
19. Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9(8):1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735
20. Huang CJ, Kuo PH (2018) A deep cnn-lstm model for particulate matter (Pm2.5) forecasting in smart cities. Sensors (Switzerland) 18(7). https://doi.org/10.3390/s18072220
21. Ibrahim H, Varol A (2020) A study on automatic speech recognition systems. 8th Int. Symp. Digit. Forensics Secur. ISDFS 2020, no. November. https://doi.org/10.1109/ISDFS49300.2020.9116286
22. Health Effects Institute (2017) State of Global Air 2017. SpecialReport. Health Effects Institute, Boston. https://www.stateofglobalair.org/sites/default/files/soga_2017_report.pdf
23. Jeya S, Sankari L (2020) Air pollution prediction by deep learning model. Proc. Int. Conf. Intell. Comput. Control Syst. ICICCS no. Iciccs, pp 736–741. https://doi.org/10.1109/ICICCS48265.2020.9120932
24. Kaya K, Gündüz Öğüdücü Ş (2020) Deep Flexible Sequential (DFS) model for air pollution forecasting. Sci Rep 10(1):3346. https://doi.org/10.1038/s41598-020-60102-6
25. Khan A, Zameer A, Jamal T, Raza A (2018) Deep belief networks based feature generation and regression for predicting wind power, pp 1–31. [Online]. Available: http://arxiv.org/abs/1807.11682. Accessed Mar 2021
26. Kostadinov S (2017) Understanding GRU Networks. https://towardsdatascience.com/understanding-gru-networks-2ef37df6c9be. Accessed Mar 2021
27. Krishana V (2018) Sparse, stacked and variational autoencoder. Medium. https://medium.com/@venkatakrishna.jonnalagadda/sparse-stacked-and-variational-autoencoder-efe5bfe73b64. Accessed Mar 2021

28. Kumar SV et al (2015) Data analysis for predicting air pollutant concentration in Smart city Uppsala. Proc 2nd ACM SIGSPATIAL Int. Work. GeoStreaming 4(6):664–672. http://uu.diva-portal.org/smash/get/diva2:919011/FULLTEXT01.pdf
29. Le VD, Bui TC, Cha SK (2020) Spatiotemporal deep learning model for citywide air pollution interpolation and prediction. Proc. - 2020 IEEE Int. Conf. Big Data Smart Comput Big Comp, pp 55–62. https://doi.org/10.1109/BigComp48618.2020.00-99
30. Lee P, Saylor R, McQueen J (2018) Air quality monitoring and forecasting. Atmosphere (Basel) 9(3). https://doi.org/10.3390/atmos9030089
31. Li X, Peng L, Hu Y, Shao J, Chi T (2016) Deep learning architecture for air quality predictions. Environ Sci Pollut Res 23(22):22408–22417. https://doi.org/10.1007/s11356-016-7812-9
32. Li L (2020) A robust deep learning approach for spatiotemporal estimation of Satellite AOD and PM2.5. Remote Sens 12(2):1–27. https://doi.org/10.3390/rs12020264
33. Liao Q, Zhu M, Wu L, Pan X, Tang X, Wang Z (2020) Deep learning for air quality forecasts: a review. Curr Pollut Rep 6(4):399–409. https://doi.org/10.1007/s40726-020-00159-z
34. Logan P. All chemistry is driven by physics. Oxford Royal Academy, Oxford. https://www.oxford-royale.com/articles/chemistry-driven-physics/. Accessed Mar 2021
35. Lv Y, Duan Y, Kang W, Li Z, Wang FY (2015) Traffic flow prediction with big data: a deep learning approach. IEEE Trans Intell Transp Syst 16(2):865–873. https://doi.org/10.1109/TITS.2014.2345663
36. Miglionico M (2019) A Deep learning framework for air pollution forecasting and interpolation. [Online]. Available: https://indigo.uic.edu/articles/thesis/A_Deep_Learning_Framework_for_Air_Pollution_Forecasting_and_Interpolation/10874069/1. Accessed Mar 2021
37. NOAA. Air Quality Research Subcommittee of the Committee on Environment and Natural NOAA, Resources CENR NOAA Aeron. Lab., no. June, 2001, [Online]. Available: https://www.esrl.noaa.gov/csd/AQRS/reports/forecasting.pdf. Accessed Mar 2021
38. Pandey VK, Rajput P (2020) Review on natural language processing. J Crit Rev 7(10):1170–1174. https://doi.org/10.31838/jcr.07.10.230
39. Qi Y, Li Q, Karimian H, Liu D (2019) A hybrid model for spatiotemporal forecasting of PM 2.5 based on graph convolutional neural network and long short-term memory. Sci Total Environ 664:1–10. https://doi.org/10.1016/j.scitotenv.2019.01.333
40. Saha S (2018) A comprehensive guide to convolutional neural networks — the ELI5 way, towards data science. https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53. Accessed Mar 2021
41. Sánchez-Balseca J, Pérez-Foguet A (2020) Spatio-temporal air pollution modelling using a compositional approach. Heliyon 6(9). https://doi.org/10.1016/j.heliyon.2020.e04794
42. Seng D, Zhang Q, Zhang X, Chen G, Chen X (2021) Spatiotemporal prediction of air quality based on LSTM neural network. Alex Eng J 60(2):2021–2032. https://doi.org/10.1016/j.aej.2020.12.009
43. Septiawan WM, Endah SN (2018) Suitable recurrent neural network for air quality prediction with backpropagation through time. 2nd Int Conf Informatics Comput Sci ICICoS 2018, pp 196–201. https://doi.org/10.1109/ICICOS.2018.8621720
44. Singh H (2018) Deep Bilief Networks- An Introduction. https://medium.com/analytics-army/deep-belief-networks-an-introduction-1d52bb867a25#id_token=eyJhbGciOiJSUzI1NiIsImtpZCI6IjEzZThkNDVhNDNjYjIyNDIxNTRjN2Y0ZGFmYWMyOTMzZmVhMjAzNzQiLCJ0eXAiOiJKV1QifQ.eyJpc3MiOiJodHRwczovL2FjY291bnRzLmdvb2dsZS5jb20iLCJuYmYiOjE. Accessed Mar 2021
45. Sharp T (2020) An introduction to Support Vector Regression (SVR). Towards data science. https://towardsdatascience.com/an-introduction-to-support-vector-regression-svr-a3ebc1672c2. Accessed Mar 2021
46. Soh PW, Chang JW, Huang JW (2018) Adaptive deep learning-based air quality prediction model using the most relevant spatial-temporal relations. IEEE Access 6(c):38186–38199. https://doi.org/10.1109/ACCESS.2018.2849820
47. Diaz G (2018) An intelligent transportation system to control air pollution and road traffic in cities integrating CEP and Colored Petri Nets, Neural Computing and Applications. https://doi.org/10.1007/s00521-018-3850-1
48. Steven NC, Walczak (2003) Artificial neural network. Science Direct. https://www.sciencedirect.com/topics/engineering/artificial-neural-network. Accessed Mar 2021
49. Sun X, Xu W, Jiang H (2019) Spatial-temporal prediction of air quality based on recurrent neural networks. Proc 52nd Hawaii Int Conf Syst Sci 6:1265–1274. https://doi.org/10.24251/hicss.2019.155
50. Thaweephol K, Wiwatwattana N (2019) Long short-term memory deep neural network model for PM2.5 forecasting in the bangkok urban area. Int Conf ICT Knowl Eng 2019-Novem:2–7. https://doi.org/10.1109/ICTKE47035.2019.8966854
51. Tsai YT, Zeng YR, Chang YS (2018) Air pollution forecasting using rnn with lstm. Proc - IEEE 16th Int Conf Dependable, Auton. Secur. Comput. IEEE 16th Int. Conf. Pervasive Intell. Comput. IEEE 4th Int.

Conf. Big Data Intell. Comput. IEEE 3rd Cyber Sci. Technol. Congr. DASC-PICom-DataCom-CyberSciTec 2018, pp 1068–1073. https://doi.org/10.1109/DASC/PiCom/DataCom/CyberSciTec.2018.00178
52. Williams R et al (2014) Air Sensor Guidebook. Epa/600/R-14/159, no 1, pp 1–5, [Online]. Available: https://cfpub.epa.gov/si/si_public_record_report.cfm?dirEntryId=277996&simpleSearch=1&searchAll=air+sensor+guidebook. Accessed Mar 2021
53. Xayasouk T, Lee HM, Lee G (2020) Air pollution prediction using long short-term memory (LSTM) and deep autoencoder (DAE) models. Sustain 12(6). https://doi.org/10.3390/su12062570
54. Xu X, Yoneda M (2021) Multitask air-quality prediction based on LSTM-autoencoder model. IEEE Trans Cybern 51(5):2577–2586. https://doi.org/10.1109/TCYB.2019.2945999
55. Yazdan A. Stacked auto encoders. https://www.researchgate.net/figure/A-stacked-autoencoder-based-deep-learning-model_fig1_310671661. Accessed Mar 2021
56. Zhang Q, Fu F, Tian R (2020) A deep learning and image-based model for air quality estimation. Sci Total Environ 724:138178. https://doi.org/10.1016/j.scitotenv.2020.138178
57. Zhang L, Liu P, Zhao L, Wang G, Zhang W, Liu J (2020) Air quality predictions with a semi-supervised bidirectional LSTM neural network. Atmos Pollut Res 12(1):328–339. https://doi.org/10.1016/j.apr.2020.09.003
58. Zhang B, Zhang H, Zhao G, Lian J (2020) Constructing a PM2.5 concentration prediction model by combining auto-encoder with Bi-LSTM neural networks. Environ Model Softw 124:104600. https://doi.org/10.1016/j.envsoft.2019.104600
59. Zhang L, Li D, Guo Q (2020) Deep learning from spatio-temporal data using orthogonal regularizaion residual CNN for air prediction. IEEE Access 8:66037–66047. https://doi.org/10.1109/ACCESS.2020.2985657
60. Zhou X, Tong W, Li L (2020) Deep learning spatiotemporal air pollution data in China using data fusion. Earth Sci Inform 13(3):859–868. https://doi.org/10.1007/s12145-020-00470-9

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Veena Gugnani received her B.Tech. Degree in Computer Science & Engineering from ITM, Gwalior. She has about sixteen years of teaching experience. She is currently pursuing the M.Tech. degree at MITS, Gwalior, Madhya Pradesh,India. Her area of interest is deep learning.

Rajeev Kumar Singh received his B.Tech. Degree in Information Technology from MGCGV, Chitrakoot, and M.Tech. degree in Software Engineering from RGPV University, India. Respectively, He has around eight years of teaching and research experience. He is actively researching soft computing methods for plant pathology. He has authored several research articles that had been published in prestigious publications and conferences. Soft computing and image processing are two of his main interests.