Forecasting of Carbon Monoxide Concentration Based on Sequence-to-Sequence Deep Learning Approach

Nur'atiah Zaini[1]([✉]) [iD], Lee Woen Ean[1] [iD], and Ali Najah Ahmed[2] [iD]

[1] Institute of Sustainable Energy (ISE), Universiti Tenaga Nasional, Kajang, Selangor, Malaysia
{Nur_Atiah,LeeWoen}@uniten.edu.my

[2] Institute of Energy Infrastructure (IEI), Universiti Tenaga Nasional, Kajang, Selangor, Malaysia
Mahfoodh@uniten.edu.my

Abstract. Carbon monoxide (CO) is one of the dangerous air pollutants due to its negative impact on human health. Therefore, accurate forecasting of CO concentration is essential to control air pollution. This study aims to forecast the concentration of CO using sequences to sequence models namely convolutional neural network and long short-term memory (CNN-LSTM) and sequence to sequence LSTM (seq2seq LSTM). The proposed forecasting models are validated using hourly air quality datasets from six monitoring stations in Selangor to forecast CO concentration at 1 h to 6 h ahead of the time horizon. The performances of proposed models are evaluated in terms of statistical equations namely root mean square error (RMSE), mean square error (MAE) and mean percentage error (MAPE). CNN-LSTM and seq2seq LSTM model excellently forecast air pollutant concentration for 6 h ahead with RMSE of 0.2899 and 0.2215, respectively. Additionally, it is found that seq2seq LSTM has slightly improved CNN-LSTM indicates the effectiveness of the architecture in the forecasting. However, both proposed architectures illustrate promising results and are reliable in the forecasting of CO concentration.

Keywords: Air quality · Forecasting · Long short-term memory · Deep learning · Artificial intelligence

1 Introduction

In recent years, air pollution has become a vital issue in most developing countries and gained worldwide attention due to its negative effects on health, economic and environmental sustainability [1, 2]. Rapid development in industrialization, infrastructure, and urbanization has caused serious air quality deterioration, especially in urban areas [3]. One of the most dangerous air pollutants namely carbon monoxide (CO) can cause negative impacts on human health such as respiratory infections, lung cancer, and heart diseases that may lead to mortality [4]. CO is a colourless, tasteless and odourless gas that is commonly emitted from the combustion of fossil fuel and coal [5]. Concentration

levels of CO are generally higher in urban areas as compared to the rural areas where the industrial, commercial and busy traffic particularly focus on the area [6]. Therefore, reliable forecasting of air pollutant concentration is essential and beneficial to provide accurate information on the air quality in the affected area and support environmental management [4].

Forecasting of time series air pollutants based on intelligent modelling strategies has been proven in illustrating higher accuracy as compared to statistical modelling such as Auto Regressive Integrated Moving Average (ARIMA) [7]. Deep learning is a subset of machine learning based on the neural network that also has been successfully implemented to solve problems in speech recognition and image classification [8]. On the other hand, deep learning strategies such as convolutional neural network (CNN) and recurrent neural network (RNN) has gained popularity in numerous studies of air quality forecasting due to their advantages over traditional machine learning models such as artificial neural network (ANN) and support vector machine (SVM) [3, 4, 9]. However, RNN is known to have a drawback during the learning process called the vanishing gradient problem [10]. Considering the limitation in RNN, an improved method namely long short-term memory (LSTM) that used memory block for recurrent learning process is introduced and have been widely applied in air quality forecasting [11, 12].

Besides that, hybrid architectures of multiple deep learning methods such as CNN-LSTM [13, 14] and sequence to sequence (seq2seq) model [15, 16] are able to improve the individual models in air quality forecasting. For instance, Wang et al. [17] developed a hybrid seq2seq model based on Bidirectional LSTM and gated recurrent unit (GRU) and Jia et al. [18] used stacked GRU layer to forecast hourly ozone concentration. Besides that, Sharma et al. [19] and Du et al. [20] developed hybrid CNN-LSTM in the forecasting of particulate matter. From the literature study, it is found that proposed hybrid architectures outperform individual deep learning models and yield the highest forecasting accuracy. However, the studies do not compare the forecasting performance between CNN-LSTM and seq2seq LSTM hybrid architectures. The comparison analysis between different hybrid models may provide new insight into the effectiveness and efficiency of hybrid architectures in air quality forecasting. Although sequence to sequence deep learning models have been previously developed for air quality forecasting, the model's evaluation in multistep forecasting of CO concentration is still limited.

The objective of this study is to establish two multistep hybrid deep learning models namely CNN-LSTM and seq2seq LSTM in hourly forecasting of CO concentration in Selangor, Malaysia. It involves hourly air quality datasets at six air quality monitoring stations for 1 to 6 h ahead forecasting of CO concentration, development and comparison of the proposed deep learning architectures. The comparison study was conducted in order to highlight the performances of different architectures and evaluate the impact of each architecture network on forecasting accuracy. The performances of the forecasting models were evaluated based on statistical evaluation such as root mean square error (RMSE), mean absolute percentage error (MAPE) and mean absolute error (MAE).

2   Data and Methods

2.1   Study Area and Data

Study Area and Data Collection. Hourly historical air quality data consist of six air pollutants namely $PM_{2.5}$, PM_{10}, SO_2, NO_2, O_3 and CO were obtained from the Department of Environment Malaysia from 1 January 2019 to 31 December 2019. The datasets were collected at six air quality monitoring stations in Selangor. Figure 1 shows the location of monitoring stations considered in this study. The hourly dataset contains 8760 records for each station. The mean hourly air pollutants concentration for the six monitoring stations are calculated and summarized in Table 1.

Fig. 1. Location of air quality monitoring stations

Table 1. Statistics of the parameters for Selangor

Variable	$PM_{2.5}$ ($\mu g/m^3$)	PM_{10} ($\mu g/m^3$)	SO_2 (ppm)	NO_2 (ppm)	O_3 (ppm)	CO (ppm)
Minimum	229.715	206.199	0.014	0.048	0.075	3.902
Maximum	7.533	3.792	0.000	0.002	0.000	0.282
Mean	41.170	31.489	0.001	0.017	0.016	0.983
Standard deviation	24.886	22.304	0.001	0.006	0.015	0.375
Total number	8760	8760	8760	8760	8760	8760

Data Preprocessing. The datasets collected contains missing values that may be due to instrumental error, invalid values and regular maintenance. In this study, mean value of the particular attribute is used to substitute the missing data. Then, mean hourly air pollutants of multiple monitoring stations were computed to represent the air quality for Selangor. The dataset was split into two sets namely training and testing. Training set is set for 80% of total records, while testing set was set to 20%. The dataset values were normalized in the range of [0, 1] to avoid the negative impacts on model's learning process due to nonuniform value ranges. The equation for data normalization is defined in Eq. 1.

$$z = \frac{x - \min(x)}{\max(x) - \min(x)} \tag{1}$$

where x is the actual value and z is the normalized value.

2.2 Long Short-Term Memory

LSTM is an updated version of RNN that is capable to learn long-term dependencies and solve vanishing gradient problems in RNN by performing self-loop memory blocks [4]. An LSTM unit consists of a memory block that includes three different gates namely forget gate, input gate and output gate as illustrated in Fig. 2. All three gates having functions of writing information from the input, forget the information, and determining the final outputs. The gate unit aims to control the information flow from one LSTM unit to another and allow the network to learn over many times steps [9].
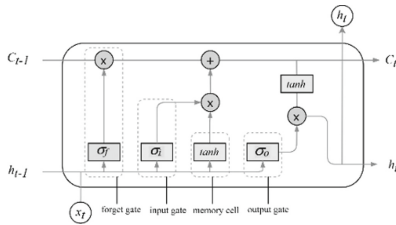
Fig. 2. LSTM unit architecture

LSTM takes current information x_t, previous output from hidden layer h_{t-1} and previous cell state, C_{t-1} as input. However, gate structures help LSTM to learn the long-term dependencies in sequential series and allow the information to pass through LSTM network. Therefore, LSTM is an effective model for learning sequential data. The forget gate, input gate, output gate and memory cell in the structure can be defined based on the following equations:

$$f_t = \sigma\left(W_f x_t + U_f h_{t-1} + b_f\right) \tag{2}$$

$$i_t = \sigma\left(W_i x_t + U_i h_{t-1} + b_i\right) \tag{3}$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \tag{4}$$

$$\tilde{C}_t = ReLU(W_C x_t + U_C h_{t-1} + b_C) \tag{5}$$

where U_f, U_i, U_o and U_C are the weight matrices connecting the preceding output to the gate units and memory cell. b_f, b_i, b_o and b_C are the bias vectors. W_f, W_i, W_o and W_C are the weight matrices mapping the hidden layer input to the gate units and a memory cell. σ denotes sigmoid function as defined in Eq. 6 and ReLU activation function is defined in Eq. 7. Then, the cell output and the layer output can be implemented using Eq. 8 and Eq. 9, respectively.

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{6}$$

$$R(z) = \max(0, z) \tag{7}$$

$$C_t = f_t \circ C_{t-1} + i_t \circ ReLU(U_C h_{t-1} + W_C x_t + b_C) \tag{8}$$

$$h_t = o_t \circ ReLU(C_t) \tag{9}$$

2.3 Convolutional Neural Network

CNN is a biologically inspired network that has been successfully implemented in image recognition, object detection and text processing [21]. CNN is also able to work on multiple arrays of data where 1D is for signals and sequences data as well as text, 2D is for images and 3D is for images taken across time and videos [22]. General CNN network architecture consists of different layers namely convolutional, max pooling, dropout and fully connected layer as illustrated in Fig. 3. In CNN, a convolutional layer is important to extract the features of input variables using the convolutional kernel [8]. The pooling layer is introduced after the convolutional layer to speed up the filtering and reduce the number of operations. Pooling layers simplifies and downsamples the output received from convolutional layers to avoid overfitting [10]. After convolutional and pooling layers, the output was flattened into 1D array for successive forecasting.
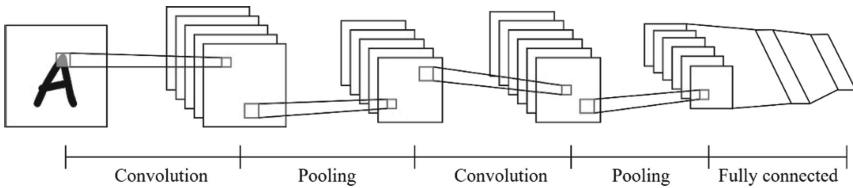

| Convolution | Pooling | Convolution | Pooling | Fully connected |

Fig. 3. CNN architecture

Considering the ability of 1D CNN in solving time series data, the application has gained worldwide attention in various fields. The equations for 1D CNN are as follows [20]:

$$c_j^l = \sum_i x_i^{l-1} * \omega_{ij}^l + b_j^l \qquad (10)$$

$$x_j^l = ReLU\left(c_j^l\right) \qquad (11)$$

$$x_j^l = Flatten\left(x_j^l\right) \qquad (12)$$

$$x_k^{l+1} = FC(\omega_{kj}^{l+1} x_j^l + b_k^{l+1} \qquad (13)$$

The convolutional layer learning process is modelled based on Eq. 10 and Eq. 11, where $*$ denotes a convolution operator, ω_{ij}^l is filter, b_j^l is bias and l is the involved layer. ReLU activation function was used within the layer. x_i^{l-1} and c_j^l represent input and output vector to a convolution layer.

2.4 Experimental Design

This study aims to evaluate the performances of two hybrid LSTM based models for CO concentration forecasting at 1 to 6 h ahead of time horizon incorporating historical air quality datasets at six air quality monitoring stations. Besides that, a comparative analysis was conducted in order to highlight the effectiveness of different hybrid architectures in forecasting 1 to 6 h ahead of CO concentration in terms of error assessments.

Seq2Seq LSTM model consists of two LSTM layers with 128 units and 64 units, respectively for both encoder and decoder processing layers. A manual search is performed to find the optimum hyperparameters of the models. The activation function used in the network is rectified linear unit (ReLU) which has the advantage of reducing the vanishing gradient and has better convergence performance. Besides that, adaptive moment estimation (ADAM) is used as an optimizer within the network where the optimizer can successfully work in online and stationary settings. The exponential decay rate for first moment estimates and second-moment estimates are 0.9 and 0.999, respectively. The learning rate is set to 0.001. Then, the forecasting models are fitted with a batch size of 128 and mean square error (MSE) is used as the loss function. Early stopping criteria is implemented for learning epoch in the model. The description of hyperparameters used in this study is summarized in Table 2.

CNN-LSTM model consists of a 1D convolution layer with a filter number of 32 and kernel size of 3. The hyperparameters of LSTM in CNN-LSTM architecture is set equal to the seq2seq LSTM model. The architecture of seq2seq LSTM and CNN-LSTM model proposed in this study are illustrated in Fig. 4 and Fig. 5, respectively.

Table 2. Hyperparameters of proposed models

Model	Hyperparameters
Seq2seq LSTM	Encoder and decoder: LSTM layer = 2, num. of nodes = {128, 64}, activation function = ReLU, learning rate = 0.001, optimizer = Adam, dropout = 0.1, batch size = 32, epoch = early stopping
CNN-LSTM	CNN: 1D convolutional layer, kernel size = 3, num. of filter = 32, activation function = ReLU LSTM: num. of layer = 2, num. of nodes = {128,64}, dropout = 0.1 Optimizer = Adam, learning rate = 0.001, batch size = 32, epoch = early stopping
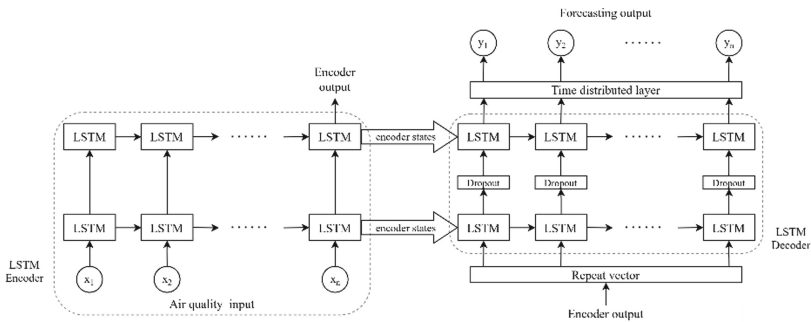
Fig. 4. Sequence to sequence LSTM architecture
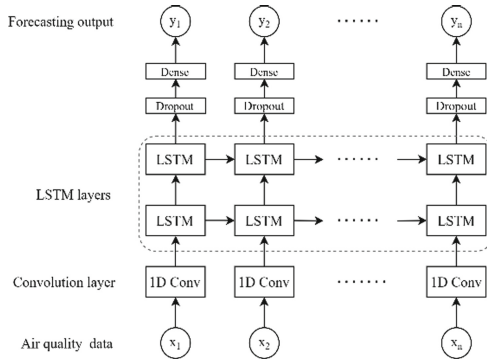
Fig. 5. CNN-LSTM architecture

2.5   Performance Evaluation

Proposed forecasting models were evaluated using statistical equations namely root mean square error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE). The RMSE represent the difference between the observed and forecasted value at a different time interval. The MAE shows the absolute difference between observed

and forecasted values on overall data points. The MAPE presents the average absolute error of forecasts in terms of percentages that measure the model's forecasting accuracy. The smaller value of RMSE, MAE and MAPE indicate better forecasting performances.

The equations of performances criteria are defined as follows:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \tag{14}$$

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{15}$$

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100 \tag{16}$$

where n is the number of data points; y_i and $\hat{y}_i$ are the observed and forecasted values, respectively.

3   Results and Discussion

The performances of CNN-LSTM and seq2seq LSTM model in the forecasting CO concentration at 1 h to 6 h ahead in terms of RMSE, MAE and MAPE are demonstrated in Fig. 6, Fig. 7 and Fig. 8, respectively. From the graphs, the error values gradually increase as the forecasting time horizon increase. It can be perceived that both forecasting models show the same trend of evaluation scores which indicate forecasting accuracy is lower for a larger forecasting time horizon [3]. Therefore, it is important to decide on the high and low resolution for optimum forecasting accuracy and reduce bias in the dataset.

The forecasting performances of proposed architectures were compared to highlight their effectiveness and impact in air quality forecasting. Both forecasting models were developed to extract input data features using the first processing layer and forecast future CO concentration using the second processing layer. In this case, encoder-decoder frameworks were proposed with different architectural designs. Seq2seq LSTM architecture yields RMSE of 0.1623, 0.1823, 0.1980, 0.2082, 0.2153 and 0.2215 for 1 h to 6 h ahead forecasting, respectively which are lower as compared to CNN-LSTM model.
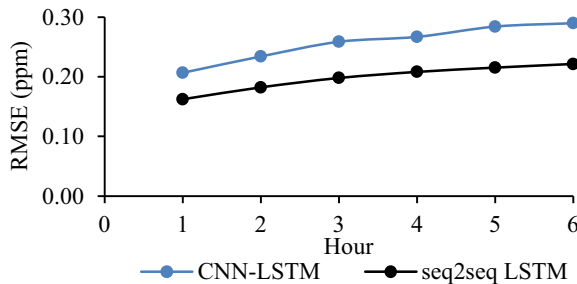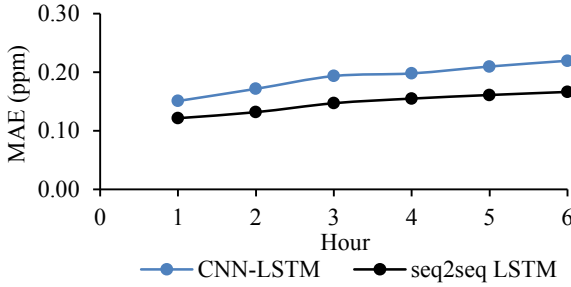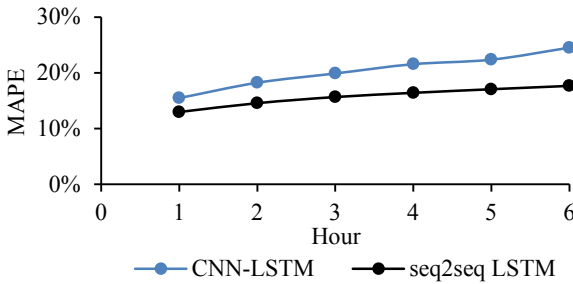

Fig. 6. RMSE of proposed models

Fig. 7. MAE of proposed models


Fig. 8. MAPE of proposed models

Similar to MAE and MAPE, the error values of seg2seq LSTM are lower than CNN-LSTM. Therefore, seq2seq LSTM model outperforms CNN-LSTM in terms of RMSE, MAE and MAPE at multi-hour step ahead forecasting.

Seq2seq LSTM reduces the RMSE, MAE and MAPE of CNN-LSTM by 23.6%, 24.2% and 28.0%, respectively at 6 h ahead forecasting. Table 3 summarizes the error values of both proposed forecasting architectures. Higher performances of seq2seq LSTM indicates the architecture successfully extracted important features and captured temporal distribution in time series air quality dataset to successfully forecast multi-hour ahead of CO concentration [15]. Therefore, the architectural design of a forecasting model affects the performances in terms of the learning process and future forecasting. However, the architecture depicts slight improvements from CNN-LSTM illustrates that CNN-LSTM may still be consistent in multi-hour CO concentration forecasting.

Overall, both CNN-LSTM and seq2seq LSTM models yield promising forecasting performances where the models are able to forecast CO concentration near the observed values. It is indicated that proposed hybrid models have the ability to extract the important features in multiple input variables and successfully forecast future CO concentration. The comparison of observed and forecasted CO concentration at 6 h ahead forecasting is presented in Fig. 9. It can be concluded that both forecasting models are reliable to forecast multistep ahead of air pollutant concentration. Different designs of architectural networks and hyperparameter combinations can be further explored to enhance forecasting performances.

Table 3. RMSE, MAE and MAPE of proposed architectures at 6 h forecasting

Model	RMSE	MAE	MAPE
CNN-LSTM	0.2899	0.2195	24.51%
Seq2Seq LSTM	0.2215	0.1663	17.66%
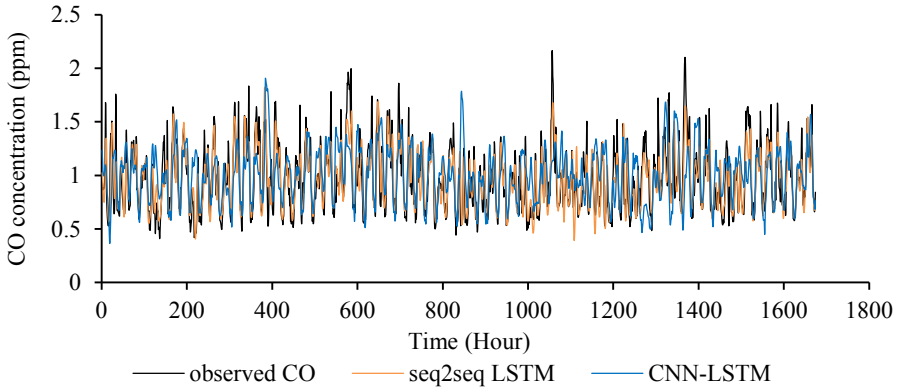
Fig. 9. Forecasted value and observed value of CO concentration

4   Conclusion

In this study, two hybrid architectures based on LSTM were proposed to forecast hourly CO concentration using air quality datasets from multiple monitoring stations in Selangor. CNN-LSTM consists of a 1D convolutional layer and two layers of LSTM. Meanwhile, seq2seq LSTM contains two LSTM layers in both decoder and decoder processing layers. Both models are designed to extract the features in multiple input variables using the first processing layer and forecast future CO concentration using the second processing layer. Seq2seq LSTM model illustrates slightly higher forecasting performances as compared to CNN-LSTM at 1 h to 6 h ahead of forecasting. However, both hybrid architectures depict superior forecasting performance and yield forecasted CO concentration near the observed values. Overall, the design of optimum hybrid architecture may depend on variational input parameters and forecasting requirements. There are many ways in which the study can be extended. First, considering other parameters such as weather and traffic data may enhance the forecasting performances which is exclusively considered in this study due to data source limitation. Second, the study can be extended by including spatiotemporal analysis among multiple air quality monitoring stations. Lastly, the hybrid architectures of deep learning approaches can be extended using more sophisticated methods such as bidirectional LSTM to handle larger datasets and optimization techniques to find optimum deep learning hyperparameters.

References

1. Ahani, K., Salari, M., Shadman, A.: An ensemble multi-step-ahead forecasting system for fine particulate matter in urban areas. J. Clean. Prod. **263**, 120983 (2020). https://doi.org/10.1016/j.jclepro.2020.120983

2. Pak, U., et al.: Deep learning-based PM2.5 prediction considering the spatiotemporal correlations: a case study of Beijing, China. Sci. Total Environ. **699**, 133561 (2020). https://doi.org/10.1016/j.scitotenv.2019.07.367

3. Chang, Y.S., Chiao, H.T., Abimannan, S., Huang, Y.P., Tsai, Y.T., Lin, K.M.: An LSTM-based aggregated model for air pollution forecasting. Atmos. Pollut. Res. **11**(8), 1451–1463 (2020). https://doi.org/10.1016/j.apr.2020.05.015

4. Zhang, B., Zhang, H., Zhao, G., Lian, J.: Constructing a PM 2.5 concentration prediction model by combining auto-encoder with Bi-LSTM neural networks. Environ. Model. Softw. **124** (2020). https://doi.org/10.1016/j.envsoft.2019.104600

5. Wong, P.-Y., et al.: Incorporating land-use regression into machine learning algorithms in estimating the spatial-temporal variation of carbon monoxide in Taiwan. Environ. Model. Softw. **139** (2021). https://doi.org/10.1016/j.envsoft.2021.104996

6. Breitner, S., et al.: Ambient carbon monoxide and daily mortality: a global time-series study in 337 cities. www.thelancet.com/. Accessed 10 May 2021

7. Liu, H., Yan, G., Duan, Z., Chen, C.: Intelligent modeling strategies for forecasting air quality time series: a review. Appl. Soft Comput. J. **102**, 106957 (2021). https://doi.org/10.1016/j.asoc.2020.106957

8. Neapolitan, R.E.: Neural Networks and Deep Learning. Springer, Heidelberg (2018). https://doi.org/10.1201/b22400-15

9. Navares, R., Aznarte, J.L.: Predicting air quality with deep learning LSTM: towards comprehensive models. Ecol. Inform. **55**, 101019 (2020). https://doi.org/10.1016/J.ECOINF.2019.101019

10. Mueller, J.P., Massaron, L.: Deep Learning for Dummies. Wiley, Hoboken (2019)

11. Yan, R., Liao, J., Yang, J., Sun, W., Nong, M., Li, F.: Multi-hour and multi-site air quality index forecasting in Beijing using CNN, LSTM, CNN-LSTM, and spatiotemporal clustering. Expert Syst. Appl. **169**, 114513 (2021). https://doi.org/10.1016/j.eswa.2020.114513

12. Rao, S., Lavanya Devi, G., Ramesh, N.: Air quality prediction in Visakhapatnam with LSTM based recurrent neural networks. Intell. Syst. Appl. **2**, 18–24 (2019). https://doi.org/10.5815/ijisa.2019.02.03

13. Li, S., Xie, G., Ren, J., Guo, L., Yang, Y., Xu, X.: Urban PM2.5 concentration prediction via attention-based CNN-LSTM. Appl. Sci. (Switzerland) (2020). https://doi.org/10.3390/app10061953

14. Huang, C.-J., Kuo, P.-H.: A deep CNN-LSTM model for particulate matter (PM2.5) forecasting in smart cities. Sensors (Switzerland) (2018). https://doi.org/10.3390/s18072220

15. Zhang, B., et al.: A novel encoder-decoder model based on read-first LSTM for air pollutant prediction. Sci. Total Environ. **765**, 144507 (2021). https://doi.org/10.1016/j.scitotenv.2020.144507

16. Du, S., Li, T., Yang, Y., Horng, S.-J.: Multivariate time series forecasting via attention-based encoder–decoder framework. Neurocomputing **388**, 269–279 (2020). https://doi.org/10.1016/j.neucom.2019.12.118,(2020)

17. Wang, H.-W., Li, X.-B., Wang, D., Zhao, J., He, H.-D., Peng, Z.-R.: Regional prediction of ground-level ozone using a hybrid sequence-to-sequence deep learning approach. J. Clean. Prod. **253**, 119841 (2020). https://doi.org/10.1016/j.jclepro.2019.119841(2020)
18. Jia, P., Cao, N., Yang, S.: Real-time hourly ozone prediction system for Yangtze River Delta area using attention based on a sequence-to-sequence model. Atmos. Environ. **244**, 117917 (2021). https://doi.org/10.1016/j.atmosenv.2020.117917
19. Sharma, E., Deo, R.C., Prasad, R., Parisi, A.V., Raj, N.: Deep air quality forecasts: suspended particulate matter modeling with convolutional neural and long short-term memory networks. IEEE Access **8**, 209503–209516 (2020). https://doi.org/10.1109/ACCESS.2020.3039002
20. Du, S., Li, T., Yang, Y., Horng, S.-J.: Deep Air quality forecasting using hybrid deep learning framework. IEEE Trans. Knowl. Data Eng. **33**(6), 2412–2424 (2021). https://doi.org/10.1109/TKDE.2019.2954510
21. Kranthi Kumar, K., Dileep Kumar, M., Samsonu, Ch., Vamshi Krishna, K.: Role of convolutional neural networks for any real time image classification, recognition and analysis. Materials Today: Proceedings (2021). https://doi.org/10.1016/j.matpr.2021.02.186
22. Lecun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature **521**(7553), 436–444 (2015). https://doi.org/10.1038/nature14539