

RESEARCH

Air quality forecasting using convolutional neural networks

G. R. Ashika* and **M. Germin Nisha**

Department of Electrical and Electronics Engineering, St. Xavier's Catholic College of Engineering, Chunkankadai, Nagercoil, India

***Correspondence:**G. R. Ashika,
ashikagr1999@gmail.com**Received:** 22 June 2023; **Accepted:** 12 August 2023; **Published:** 26 September 2023

Air pollution is now one of the biggest environmental risks, which causes more than 6 million premature deaths each year from heart diseases, stroke, diabetes, respiratory disease, and so on. Protecting humans from the damage which is caused by air pollution is one of the major issues for the global community. The prediction of air pollution can be done by machine learning (ML) algorithms. ML combines statistics and computer science to maximize the prediction power. ML can be also used to predict the air quality index (AQI). The aim of this research is to develop a convolutional neural network (CNN) model to predict air quality from the unseen data set, which includes concentration of nitrogen dioxide (NO₂), carbon monoxide (CO), and sulfur dioxide (SO₂). The proposed system will be implemented in two steps; the first step will focus on data analysis and pre-processing, including filtering, feature extraction, constructing convolutional neural network layers, and optimizing the parameters of each layer, while the second step is used to evaluate its model accuracy. The output is predicted as AQI for the developed CNN model. The developed CNN model achieves a root-mean-square error of 13.4150 and a high accuracy of 86.585%. The overall model is implemented using MATLAB software.

Keywords: air quality prediction, air pollution, deep learning algorithm, convolutional neural network, air quality index

1. Introduction

Artificial intelligence (AI) is a powerful tool that is used for measuring and solving people's problems. AI technology allows machines to mimic human behavior. The different neural network architectures like deep neural network (DNN), recurrent neural network (RNN), and convolutional neural network (CNN) and machine learning algorithms ensure that future air quality indexes (AQIs) can be predicted. Supervised learning, unsupervised learning, and reinforcement learning are the learning algorithms used in machine learning (ML) of AI (1).

Air pollution refers to indoor or outdoor air pollution. When pollutants enter the environment, the air becomes polluted, making plants, animals, and humans unable to live. One of the leading causes of death is air pollution (2). Air pollution killed 6.6 million people in 2020. Therefore, we need a reliable forecasting system. Sulfur dioxide (SO₂), carbon monoxide (CO), particulate matter (PM), nitrogen

oxide (NO), and other pollutants are the most harmful air pollutants. AQI is a standard used to measure the air quality.

This article focuses on an important aspect of deep learning such as CNN that is used to predict quality of air. The aim of this research is to develop a CNN model to predict air quality from invisible data, including the concentration of pollutants.

2. Air quality prediction model

The proposed air quality prediction (AQP) model is mainly based on CNN.

2.1. Deep learning algorithm

Deep learning uses ANNs to perform the complex calculations on large datasets. It is based on the structure

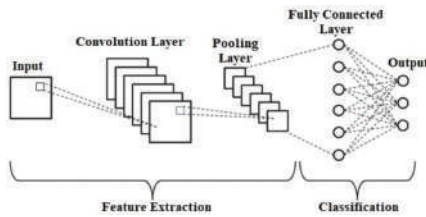


FIGURE 1 | Architecture of convolutional neural network (CNN).

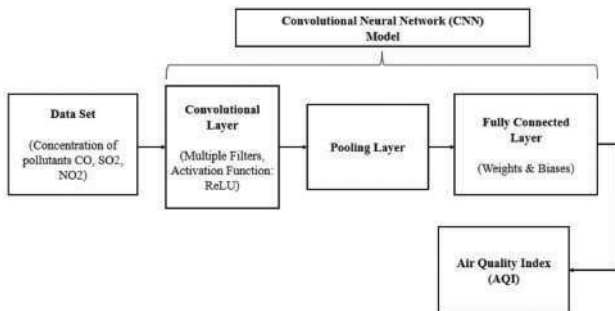


FIGURE 2 | Proposed System Block Diagram.

and function of the human brain. This algorithm trains the machines by learning it from the examples. Deep learning architectures are designed to solve the problem of climate change (3). Deep learning is performed by DNNs. There are many different DNN architectures such as CNNs. It is also classified according to whether it is learned with supervision or not. Unsupervised DNNs can use unlabeled data, while supervised DNNs must collect training data for the training. The CNN we use in this article is a supervised DNN.

2.2. Convolutional neural network

Convolutional neural network (CNN) is a neural network architecture used for deep learning. The structure of CNN has a three-layer architecture shown in Figure 1, where the first layer is usually called a convolutional layer, the next layer is called a pooling layer, and the final layer is a fully connected layer. It is used in many applications like image classification (4), face recognition, object detection, and so on. They are also great for audio classification and distribution, time-series, and signal data.

2.2.1. Convolutional layer

This layer is the main building block of CNNs. It has a set of filters whose parameters will be examined throughout the training process. The filter size is usually smaller than the actual image. Each filter intertwines with the image and creates an activation map.

2.2.2. Pooling layer

The pooling layer is mainly responsible for reducing the spatial size of the convolved feature. Pooling layers down

sample the feature maps by reducing their size. The two types of pooling (5) are: (i) **Max pooling**: the maximum pixel value of the selected batch. (ii) **Average pooling**: the average of all pixels in a group is selected.

2.2.3. Fully connected layer

The FC layer is a neural network in which each neuron applies a linear transformation from a weight matrix to an input vector. The FC layer has weights, biases, and neurons. An FC layer takes and multiplies the input by the weight matrix and then it adds the bias vector.

3. Proposed CNN model

The proposed model is based on CNN. The system uses CNN supervised deep learning algorithm to monitor the air quality. The deep learning architecture is designed to solve air pollution prediction problems, climate change problems, non-linear, seasonal, cyclical, and sequential dependency problems between the pollutant data.

Figure 2 represents the block diagram for predicting the quality of air using a CNN.

In Figure 2, data sets of sulfur dioxide (SO_2), nitrogen dioxide (NO_2), and carbon monoxide (CO) are given as input. A convolutional layer consists of a set of filters that will filter the input. In CNN, the convolutional layer extracts various features of the input. The output is termed as a feature map. Next, the pooling layer will reduce the size of the feature map. Therefore, the number of computations in the network is reduced and the computational costs are also reduced. An FC layer consists of weights, biases, and neurons and is used to connect neurons in two different layers. The AQI will be predicted for the developed CNN model.

4. Air quality index

Air quality index (AQI) is used to indicate everyday air quality (6). The index's daily values are used to communicate air pollution forecasts to the public. An increase in the AQI means an increase in air pollution that threatens human health. The AQI ranges from zero to five hundred (0-500) (7). The composition of each pollutant is different; that is why AQI values are categorized for public health warnings and color codes.

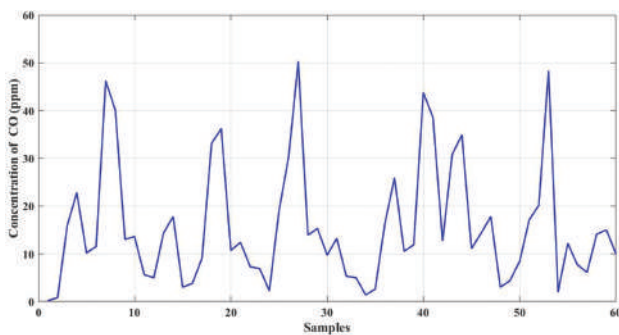
1. 0-50 indicates AQI is "Good"
2. 51-100 indicates AQI is "Moderate"
3. 101-150 indicates AQI is "Unhealthy for some Sensitive Groups"
4. 151-200 indicates AQI is "Unhealthy"
5. 201-300 indicates AQI is "Very Unhealthy"

TABLE 1 | Breakpoints for air quality index (AQI).

AQI categories (Index Values)	CO (ppm)	SO ₂ (ppb)	NO ₂ (ppb)
Good (0–50)	0–4.4	0–35	0–53
Moderate (51–100)	4.5–9.4	36–75	54–100
Unhealthy for some sensitive groups (101–150)	9.5–12.4	76–185	101–360
Unhealthy (151–200)	12.5–15.4	186–304	361–649
Very unhealthy (201–300)	15.5–30.4	305–604	650–1249
Hazardous (301–500)	30.5–50.4	605–1004	1250–2049

TABLE 2 | Data set samples.

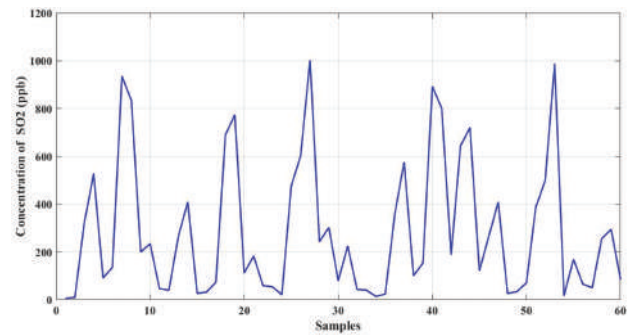
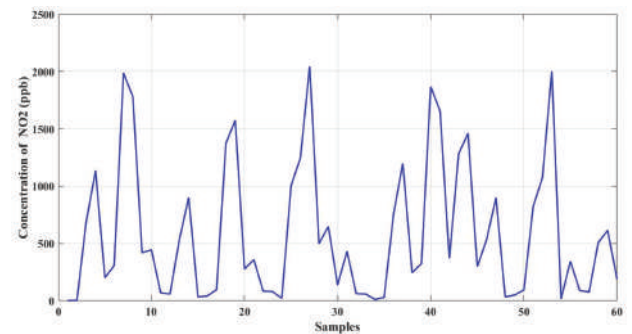
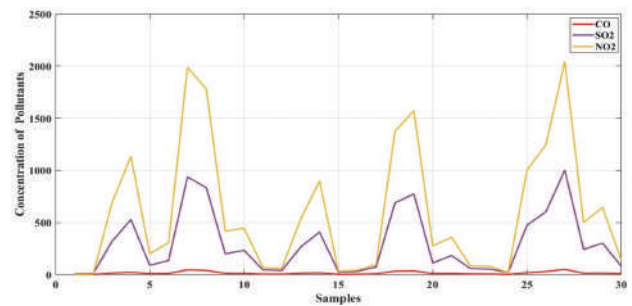
CO (ppm)	SO ₂ (ppb)	NO ₂ (ppb)	AQI
0.2	6	0.8	1
0.8	10	3	5
15.9	320	691	205
22.8	528	1,134	287
10.2	91	201	112
11.5	136	307	134
46.2	936	1,987	447
40	834	1,780	415
13	200	418	159
13.6	234	444	167

**FIGURE 3** | Concentration of CO samples.

6. Greater than 300 indicates AQI is “Hazardous”

The AQI is calculated by determining the concentration of pollutants from a linear function.

Table 1 shows the AQI breakpoints for three pollutants: sulfur dioxide (SO₂), carbon monoxide (CO), and nitrogen dioxide (NO₂). These breakpoints are used to identify health practices for each AQI group, so people can understand their health impacts and protect themselves from them.

**FIGURE 4** | Concentration of SO₂ samples.**FIGURE 5** | Concentration of NO₂ samples.**FIGURE 6** | Training input.

5. Results and discussion

Here, some of the experimental results are presented to demonstrate the effectiveness of the proposed prediction model. The data set of the concentrations of pollutants CO, SO₂, and NO₂ are used in the experiments.

5.1. Datasets

The sample data set for developing the CNN model is represented in **Table 2**. It shows the sample data sets of the pollutants nitrogen dioxide (NO₂), sulfur dioxide (SO₂), and carbon monoxide (CO) and their corresponding AQI values for developing the CNN model. These data sets are trained and tested for the development of the CNN model.

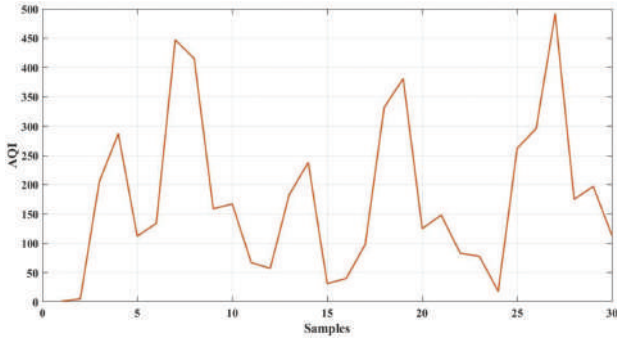


FIGURE 7 | Training output.

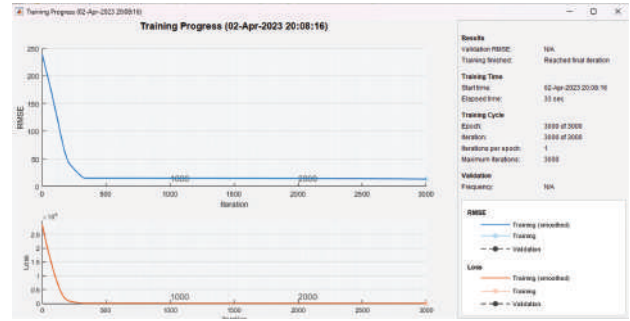


FIGURE 10 | Training progress.

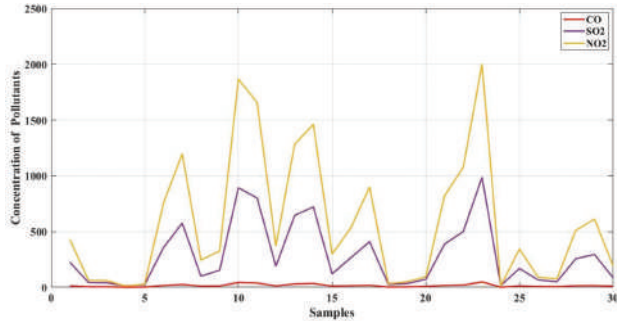


FIGURE 8 | Testing data.

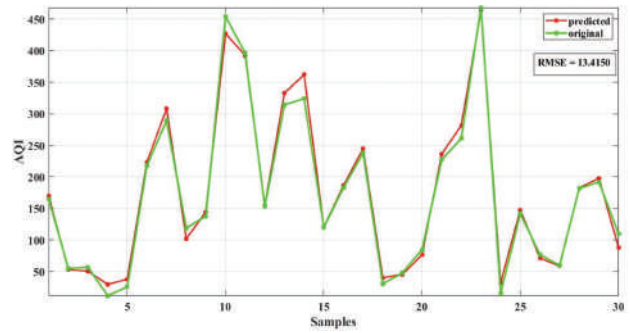


FIGURE 11 | Original and predicted output of convolutional neural network (CNN) model.

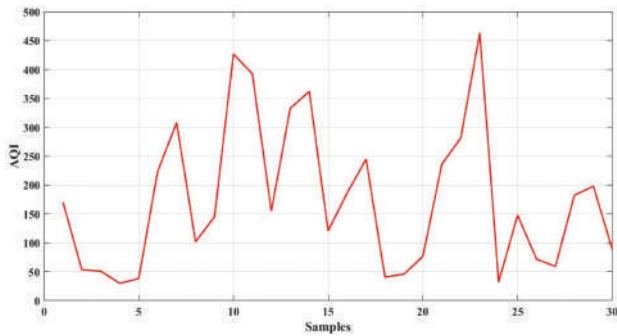


FIGURE 9 | Predicted output.

5.2. Concentration of pollutants

The graphical representation of data sets for concentrations of pollutants nitrogen dioxide (NO₂), sulfur dioxide (SO₂), and carbon monoxide (CO) are represented.

Figure 3 shows the graphical representation of concentration of CO samples in the data sets. The concentration of CO is trained and also tested for the development of the future CNN model. The range of concentrations of CO samples which are taken for the development of the CNN model is 0-50.4 (ppm).

Figure 4 shows the graphical representation of concentration of SO₂ samples in the data sets. The concentration of SO₂ is trained and also tested for the development of the upcoming CNN model. The range of

concentrations of SO₂ samples which are taken for the development of the CNN model is 0-1004 (ppb).

Figure 5 shows the graphical representation of concentration of NO₂ samples in the data sets. The concentration of NO₂ is trained and also tested for the development of the future CNN model. The range of concentrations of NO₂ samples which are taken for the development of the CNN model is 0-2049 (ppb).

5.3. Training and testing data

Figure 6 shows the graphical representation of training input taken from the data sets. The concentrations of the pollutants CO, SO₂, and NO₂ are taken as training input for the development of the CNN model.

Figure 7 shows the graphical representation of training output taken from the data sets. The AQI values are taken as training output for the development of the CNN model.

Figure 8 shows the graphical representation of testing data taken from the data sets. The concentrations of the pollutants CO, SO₂, and NO₂ are taken as testing input for the development of the CNN model.

5.4 Predicted output

Figure 9 shows the graphical representation of predicted output. The AQI values are obtained.

Figure 10 shows the graphical representation of the training progress of the predicted output of the developed CNN model.

Figure 11 represents the original and predicted output of the developed CNN model. AQI is obtained as the result.

The root mean square error (RMSE) obtained for the developed CNN model is 13.4150.

6. Conclusion

An effective air quality forecast model based on the CNN with supervised deep learning algorithm is proposed in this article. The overall system is implemented by the MATLAB software. The data set samples are collected. The collected sample data sets are trained and then tested for development of the CNN model. Then from the collected sample data set, the training inputs are converted into 4D arrays for training. The CNN is constructed with all layers. After that, the training process will be done for the CNN. Then the testing is done, and the testing data from the sample data set are converted into 4D arrays for the testing process. The testing process is done for the CNN model. The output is predicted for the developed CNN model as the AQI. The developed CNN model achieves RMSE of 13.4150 and 86.585% accuracy. In future, more efficient suitable deep learning models can be developed to predict air quality.

Author contributions

Both authors made important contributions to developing the CNN model, which has data analysis and pre-processing and also contributed in testing the accuracy of the developed model using unseen datasets.

References

1. Aditya CR, Vidyavastu P, Nayana DK, et al. ML models for air pollution detection and prediction. *Int J Eng Trends Technol*. 59:204–7.
2. Maleki H, Goudarzi G, Rahmati M, Baboli Z, Sorooshian A, Birgani YT. Air pollution prediction using an artificial neural network model. *Clean Technol Environ Policy*. (2019) 21:1341–52.
3. Horng S, Li T, Yang Y, Du S. The use of a hybrid deep learning framework for deep air quality forecasting. *IEEE Trans Knowl Data Eng*. (2021) 33:2412–24.
4. Zhang J, Xie Y, Rijal N, Yang W, Bo Q. Utilising convolutional neural networks and weather features for particle pollution estimation from images. *Proceedings of the 25th IEEE International Conference on the Image Processing, Athens, Greece*. Athens: (2018). p. 3433–7.
5. Chauhan R, Kaur H, Alankar B. Convolutional neural networks are used for air quality forecasting in urban environments to promote sustainable development. *Sustain Urban Commun Soc*. (2021) 75:103239.
6. Bedekar M, Singh VM, Barve A, Shirao S. LSTM cells and parallel dense neural networks for air quality index forecast. *Proceedings of the 2020 International Conference for Emerging Technology (INCET), Belgaum, India*. Belgaum: (2020).
7. Yu D, Gu Y, Liu H, Li Q. Using machine learning algorithms, the air quality index and air pollutant concentration may be predicted. *Appl Sci*. (2019) 9:4069.