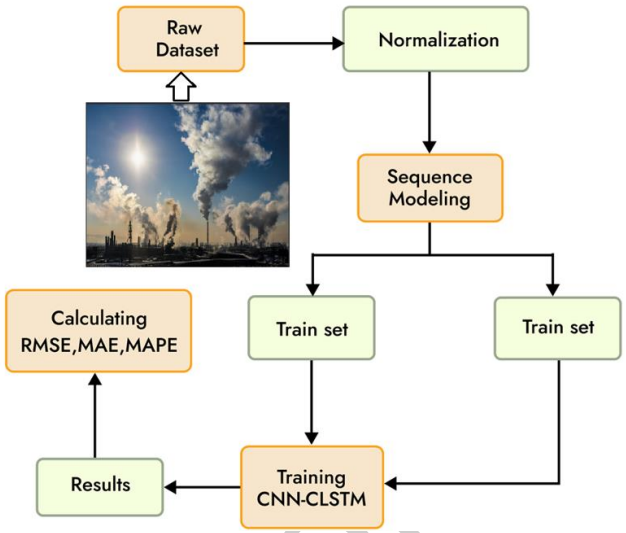
Global NEST Journal, Vol 27, No X, pp 1-13

Ozone concentration forecasting using a smart hybridization of deep learning models with real-time data

Gopal Rathinam[1], Jaya Sinha[2], Chandraprabha Muthukrishnan[3], Vidhyavathi kotha[4], P.Preethi[5]* and R.Aruna[5]
[1]Information and communication Engineering, College of Engineering, University of Buraimi, Al Buraimi, Oman
[2]Department of CSE & CSE-AIML, ITS Engineering College, Greater Noida, India
[3]Department of AI & DS, GITAM University, Bengaluru, India
[4]Deptartment of CSE, CMR Institute of Technology, Bangalore, India
[5]Associate Professor, Department of Information Technology, Kongunadu College of Engineering and Technology, Trichy.
Received: 23/05/2025, Accepted: 22/07/2025, Available online: 23/07/2025
*to whom all correspondence should be addressed: e-mail: preethi1.infotech@gmail.com
https://doi.org/10.30955/gnj.07683

Graphical abstract



Abstract

Surface ozone is a critical air pollutant responsible for significant public health impacts across global cities, with more than 365,000 premature deaths in 2019. Good forecasting of the concentrations of surface ozone is important for effective public health measures and management of air quality standards. However, current methods face challenges in dealing with the complexity of meteorological dynamics and spatial variability sometimes leading generally to uncertain predictions. Considering these aspects, we design a hybrid CNN and Concatenated CLSTM in regard to modeling the improvement of surface ozone forecasting: the hybrid CNN-CCLSTM model. It uses spatial learning of CNN to extract spatial features and convoluted temporal dependencies of Concatenated CLSTM to better account for variability in meteorological conditions. In this model, an ensemble forecasting approach is used to represent uncertainty in the climate patterns as well as their projection for the weather. To prove the efficiency of the proposed CNN-CCLSTM model, we conducted the experiment over Delhi city in the region of India, comparing predictions for two seasons of pollution. The results of this process show that such a hybrid model does improve the accuracy of daily predictions of ground-level ozone concentrations and hence allows effective measurement of the associated uncertainties caused by variability in the weather. This, in turn, provides a more reliable tool for air quality management and public health protection.

Keywords: Surface Ozone Forecasting, Hybrid CNN-CCLSTM Model, Air Quality, Meteorological Variability, Ensemble Forecasting and Climate Models.

1. Introduction

Surface ozone is a major warm-season air pollutant in global cities (MEEC, 2021; USEPA, 2021), causing 365,000 premature deaths worldwide in 2019 [Murray *et al.* 2020]. Many cities have implemented evacuation strategies in response to the health risks posed by ozone depletion. These strategies include reducing precursor greenhouse gases or issuing public health advisories when surface ozone concentrations are expected to be higher than the local air quality norms (People's Republic of China Ministry of Environment, 2015; USEPA, 2015).

Predicting ozone levels many days beforehand is crucial for efficient emergency reaction procedures. MEEC mandates daily city-level air quality predictions for the next 5 days, with ≥60% accuracy expected for the next 1–3 days (MEEC, 2020). Furthermore, the Maximum Daily 8-hour average (MDA8) levels of ozone forecasts must have an accuracy of ±10 for non-exceedance days and ±15 for light pollution days. This translates to ±12 and ±16.5 μg m−3 for the Air Quality Index (AQI) projections, or ±8 and ±30 μg m−3 of per hour ozone concentration projections, correspondingly). Nonetheless, an earlier investigation revealed that the RMSE of hourly estimates for ozone concentration in 34 Chinese cities came from seven.

Using operable 72-hour forecast systems, summertime concentrations were around 40 μg m−3, with regional

variations [Petersen *et al.* 2019], occasionally meeting the accuracy standards set by the MEEC. What elements restrict surface accuracy ozone projections, and if the present prediction technologies can provide the required forecast accuracy levels, have not undergone a thorough assessment.

The photochemical synthesis of oxygen and the buildup of its precursors are facilitated by certain weather circumstances, such as elevated temperatures, intense sun radiation, and surfaces converging winds and stagnant boundary layers [Fu *et al.* 2019; Jacob and Winner 2009]. Different synoptic to mesoscale weather patterns are responsible for these local circumstances. The Pearl River Delta (PRD) in Southern China regularly has ozone pollution incidents due to subsidence or surface wind confluence brought about by the west Pacific continental high, which is the outflow of an advancing storm [Li *et al.* 2022; Ouyang *et al.* 2022] or the wind between the land and the sea [Ding *et al.* 2004]. Therefore, the Synoptic (typhoon track) to mesoscale (land-sea breeze) weather predictions provide significant variability to surface ozone projections [Preethi *et al.* 2024; Wang *et al.* 2021]. But as of right now, because they rely on a limited number of regional air quality simulations, operational air quality predictions are unable to capture the whole spectrum of potential future meteorological conditions. Furthermore, lead increases the uncertainty of predictions for the weather time (the interval of time preceding the release of a prediction and the anticipated event's occurrence) [Slingo and Palmer 2011] and might put air quality estimates above a certain point in terms of "certainty."Lead time could not be precise enough to meet specifications or management demands. The non-linearity associated with atmospheric dynamics and its inherent uncertainties are the primary causes of forecasting uncertainties innate sensitivity to beginning circumstances, or the ability to detect even little changes in the original conditions [Sathyasri *et al.* 2019] would cause a significant and expanding divergence in the functioning of the system. In contemporary techniques for predicting the weather, forecast an ensemble of thirty to fifty forecasting members, each of whom is a model implementation with little changes to the original circumstances or physical characteristics, allowing the individuals in the ensemble to represent a variety of potential weather scenarios. It is sense that surface ozone estimates' meteorological variability should also be measured using an ensemble method. However, the computational cost of 3-D metropolitan air quality modeling is high such that it is not practical for everyday operations to run several simulations with erratic weather predictions.

Here, we suggest using ML/DL techniques to effectively perform surface ozone ensemble projections and measure the degree of uncertainty in the weather. Prior research has indicated the success of ML/DL techniques in air quality predictions [Athira *et al.* 2018; Sayeed *et al.* 2020; Sun *et al.* 2021]. But prior ML/DL models were mostly trained using locally reported weather parameters and pollution levels, which eliminated the spatial information that was continental to mesoscale flow. Meteorological and pollutant observations made during urgent pollution control activities may have an impact. Additionally, China and other emerging nations have seen fast changes in precursor emissions, to the point where the ML/DL historical measurement models may not accurately represent the antecedent force that exists now.

We integrated the 2DCNNs approach, which focused on spatial trends [Huang *et al.* 2021; Xing *et al.* 2020] using weather ensemble projections to reconstruct surface ozone. The objective of the ensemble forecasting method (2DCNN-SOEF) is to measure the climatic uncertainty of ozone predictions. We used a broad range of perturbations to a region air quality models to create a sizable training information set variety of weather variables from continental to mesoscale scale. To demonstrate the notion, we built a daily 216-hour China's PRD region's Shenzhen City uses the 2DCNN-SOEF technology. We compared measurements from two polluting seasons to assess the prediction uncertainty and skill of the 2DCNN-SOEF method. The novel contributions are,

The proposed hybrid CNN-CCLSTM model captures the spatial and temporal dynamics more effectively, thus improving the accuracy of the surface ozone forecast. This advance has important implications for making increasingly reliable predictions about air quality, which is crucial for public health initiatives and compliance with requirements.

This makes the approach systematic while quantifying and analyzing uncertainties that arise from ozone predictions in the model to contingent variable meteorological conditions. This capability may improve understanding of factors influencing air quality and risk assessment.

The hybrid model has a design that suits the changes in precursor emissions speeds, especially in growing economies. This ensures that even as conditions change in the environment and the regulatory framework, predictions remain relevant and accurate and therefore contributes to better management of air quality.

2. Materials and methods

2.1. Data collection and pre-processing

This study applied a model that was developed based on climatic and air quality measurements collected over the period of January 2015 to December 2017 in Delhi, India. In general, climate of Delhi is classified as continental and contains enormous seasonal variations. The average annual temperature in Delhi is about 25.0 °C with seasonal and yearly fluctuations of -1.0 °C in January and +45.0 °C in July. The summers are dry-hot, and winters cool-foggy, while spring and autumn seasons are short. Precipitation is unevenly distributed; at between June and September, the monsoon season accounts for about 80% of annual rainfall. The Delhi region has a wide net of meteorological monitoring stations and air quality measuring sites that have fairly wide coverage across

urban and suburban areas. **Figure 1** Location of pollution and meteorological observation facilities in Delhi There were two main selections that were considered in the selection of input data: first, that the surveillance data from the monitoring stations be as close as possible to the central area of Delhi; second, that all sets of surveillance data contain both air quality and meteorological monitoring locations as near to each other as possible.

Under the above conditions, the average value of the data from 15 pairs of observation sites was used as input data to our ozone prediction model. Each pair of environmental and climatic monitoring stations is separated by a distance of not more than 3 kilometres. This configuration provided air quality of good quality over Delhi, while the meteorological data were sourced from the India Meteorological Department (IMD) and National Environmental Engineering Research Institute (NEERI). **Table 1** shows atmospheric and air quality data used for this study. Short data gaps (<3 h) are linearly interpolated, while longer or monsoon-related outages are imputed using a k-nearest-neighbour approach (k = 5) before normalisation and trend decomposition.

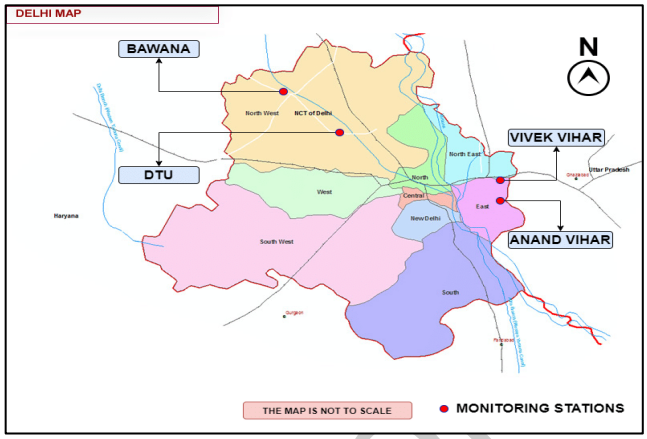


Figure 1. The locations of Delhi City's weather and atmospheric surveillance stations [Tiwari *et al.* 2021]

Table 1. Data on air quality and weather were included into the modeling analysis.

Category	Feature Name	Description	Scale/Unit	Period	Data Source
Air Quality	AQI	Composite air quality index	Integer (0–500)	Daily	CPCB (Central Pollution Control Board)
	AQI Level	AQI category (e.g., Good, Poor, Severe)	Categorical (6 levels)	Daily	CPCB
	PM2.5	Particulate matter ≤2.5 µm	µg/m³	Daily	CPCB
	PM10	Particulate matter ≤10 µm	µg/m³	Daily	CPCB
	SO_2	Sulfur dioxide concentration	µg/m³	Daily	CPCB
	CO	Carbon monoxide concentration	mg/m³	Daily	CPCB
	NO_2	Nitrogen dioxide concentration	µg/m³	Daily	CPCB
	O_3 (8-hour mean)	Average ozone over 8 hours	µg/m³	Daily	CPCB
Meteorological	Max Temperature	Daily maximum temperature	°C	Daily	IMD (Indian Meteorological Department)
	Average Temperature	Daily average temperature	°C	Daily	IMD
	Daytime Climate Grade	Weather rating during 6 AM–6 PM	Ordinal scale (1–5)*	Daily	Derived from IMD data
	Evening Climate Grade	Weather rating during 6 PM–12 AM	Ordinal scale (1–5)*	Daily	Derived from IMD data
	Wind Direction	Dominant wind direction	Compass (e.g., N, NW)	Daily	IMD
	Wind Speed Grade	Wind speed classification	Ordinal scale (1–5)**	Daily	IMD
Target Variable	Avg. Ozone Level (Next 8 Days)	Predicted ozone concentration (8-day average, 8-h mean)	µg/m³	8-day forward window	Forecast Output

If Xr is the name given to the gathered data, then Xr may be symbolized by

$$\mathrm{Xr} = \left[\left[\mathrm{Xr1(t)}\right]^{\mathrm{T}}\left[\mathrm{Xr2(t)}\right]^{\mathrm{T}}\ldots\left[\mathrm{Xrm(t)}\right]^{\mathrm{T}}\right] \quad (1)$$

$$= \begin{bmatrix} Xr1(1) & Xr2(1) & Xrm(1) \\ Xr1(2) & Xr2(2) & Xrm(2) \\ Xr1(T) & Xr2(T) & Xrm(T) \end{bmatrix} \quad (2)$$

where M and T stand for the total amount of input characteristics and variables inside relationships, and the value of $Xr_i(j)$ is the ith input feature of the jth day. The gathered information was normalized as follows according to the distributional properties of

$$Xn_i(t) = \frac{Xr_i(t) - Xr_i(t)_{mean}}{Xr_i(t)_{std}} \quad (i = 1,2,\ldots,M) \quad (3)$$

$$Xn = [\left[Xn1(t)\right]^T\left[Xn2(t)\right]^T \ldots\ldots\left[XnM(t)\right]^T \quad (4)$$

where $Xr_i(t)$ mean and $Xr_i(t)_{std}$ are the logical expectations and standard deviation of $Xr_i(t)$, each of which Xn is the standard deviation of the data. As seen in **Figure 2**, the starting point X of the CNNCLSTM built from

normalized information X_n is expressed by the following solution.

$$X = \left[X_P X_{P+1} \ldots . X_T\right] \quad (5)$$

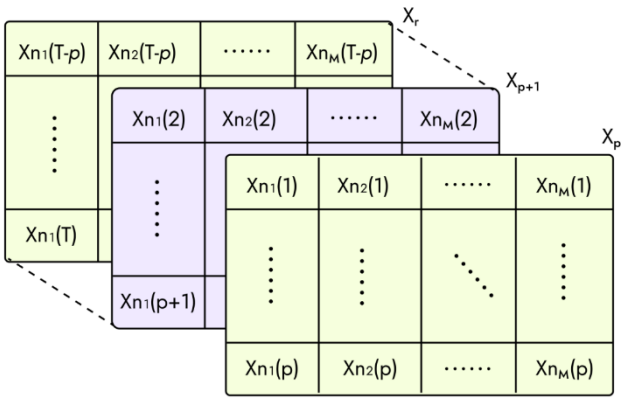


Figure 2. The CNN-CLSTM algorithm's processed input information layout

3. Proposed CNN-CLSTM hybrid model

3.1. CNN model taxonomy

CNN is a highly helpful neural network that was developed from human brain systems and performs very well in a variety of applications. CNN is characterized by shared weight and patchy connection. Using a sliding-window technique, the layers of convolution gathers characteristics of the input information in order to create mappings of features that represent the temporal organization characteristic of time series information. Feature translation is performed via a convolutional filter with weight distributed all through the layer and directly connected to input information. Subsampling processes decreases result dimensionality by averaging or maximally pooling mappings of features in the convolutional layer, allowing for the ignoring of small shifts or distortions in the information being supplied. The output information of the CNN algorithm is produced by the last fully linked layer of CNN. **Figure 3** depicts CNN analysis for multidimensional period information, where M and N represent input characteristics and time durations, and C_i and S_i represent the i^{th} convoluted and subsampling levels. Let C_{Ki} and C_{Ni} stand for the kernel size and the total time duration in the C_i layer, respectively. Similarly, SPi and SNi stand for the total pooling area and duration duration in the Si layer. $CN1 = N - CK1 + 1$ follows, and in a similar manner, $CN_2 = SN_1 - CK2 + 1$, and $SNi = C_{Ni}\, S_{Pi}$. The terms CMi and SMi refer to the number of filter outputs in the Ci and Si layers, respectively. The result of an entirely connected level is the result of the network's components as a whole. Convolutional and subsampling processes layers are therefore alternated while CNN conducts training experiments ith information in order to completely capture the properties of periodic or sequential information.

3.2. CLSTM models

CLSTM is a short-term memory to address the long-term dependence issue by learning dependencies that last. Recognition of speech, processing of natural languages, and automated picture labeling are just a few of the sequence information challenges that are being solved by CLSTM since it can handle even the longest piece of information without the gradient disappearing. An intricate recurrent architecture inside a single cell of the CLSTM is seen in **Figure 4** and is related historically throughout time. CLSTM has two property principles: the cell's hidden state H(t) evolves over time, and the cell state C(t) maintains memory over time. **Figure 4** shows an upward shift in the cell state across the highest line of the blocks diagram's CLSTM cell. The data contained in the state of the cell may be added or removed by the CLSTM. The forget gate F(t) alters the link between input X(t) and prior hidden state H(t-1) to cell state C(t), either remembering or forgetting them as required. Determining the cell condition C(t), the output gates O(t) also chooses the exit. In the proposed concatenated CLSTM (CCLSTM) architecture, convolutional feature maps are concatenated with each LSTM hidden state at every time step, enabling unified spatial–temporal representation learning.

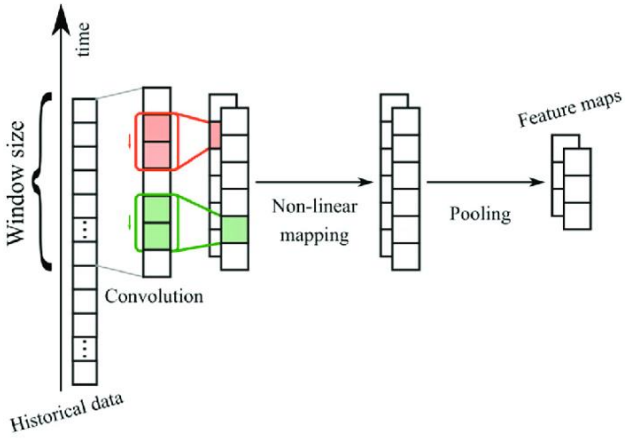


Figure 3. CNN architecture for information derived from multivariate time periods

Forget Gate:

$$f_t = \sigma\left(W_f \cdot \left[H_{t-1}, X_t\right] + b_f\right) \quad (6)$$

Input Gate:

$$i_t = \sigma\left(W_i \cdot \left[H_{t-1}, X_t\right] + b_i\right) \quad (7)$$

Output Gate:

$$o_t = \sigma\left(W_o \cdot \left[H_{t-1}, X_t\right] + b_o\right) \quad (8)$$

Candidate Cell State:

$$\tilde{C}_t = \tanh\left(W_c \cdot \left[H_{t-1}, X_t\right] + b_c\right) \quad (9)$$

Cell State Update:

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \quad (10)$$

Hidden State Update:

$$H_t = o_t \odot \tanh\left(C_t\right) \quad (11)$$

Equations (6) through (11) represent the internal operations of the standard LSTM unit used within our

CLSTM architecture. The input X_t and previous hidden state H_{t-1} are concatenated and passed through three sigmoid-activated gates: forget (f_t), input (i_t), and output (o_t). These gates control the flow of information through time. The candidate cell state $\tilde{C}_t$ is generated using a hyperbolic tangent function. The final cell state C_t is updated using a combination of the previous memory and new information, while the hidden state H_t is modulated by the output gate. The operator $\odot$ denotes element-wise multiplication, ensuring selective memory retention and update across time steps. This formulation supports robust temporal learning in ozone forecasting tasks. W and B represent the weight matrices and bias vectors, respectively; σ(·) represents a sigmoid function, and tanh(·) represents the function with hyperbolic tangent. There is close oversight over the CLSTM's internal configuration and input information, which can be seen in the current state of the cell because of how it works. When CLSTM is used in conjunction with other kinds of deep neural networks, these benefits become more apparent than when CLSTM is used alone. Robustness is enhanced through a bootstrap ensemble of 30 CCLSTM models trained on resampled subsets, whose forecasts are combined via validation-error-weighted averaging.

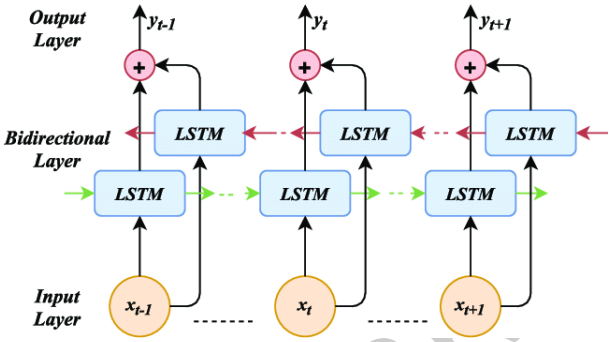


Figure 4. CLSTMs framework for Network

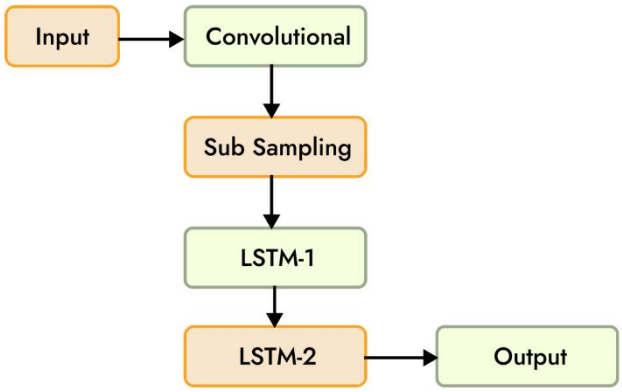


Figure 5. CNN-CLSTM hybrid design

3.3. CNN-CLSTM integration phase

In terms of ozone forecasting, the aforementioned CNN and CLSTM algorithms offer benefits and drawbacks, respectively. Given the size of the kernel, CNN may extract an extensive list of helpful characteristics to guarantee precise forecasting from the input time series information. Put differently, CNN may determine whether characteristics represent a longer or narrow time period from multimodal time series information by adjusting the kernel size. In contrast, CNN is better at capturing the critical input information properties required for forecasting, whereas CLSTM is less effective at reflecting the long-term past processes in the input time series information. This combined approach effectively complements the two distinct approaches mentioned above. This work uses CNN to derive the vectors of features from multimodal time series information for ozone forecast. The CLSTM layers then uses these findings to train the model for forecasting. The combination of models that combines CNN and CLSTM is seen in **Figure 5**.

3.4. Training and testing of CNN-CLSTM hybrid model-based ozone predictor

Figure 6 displays the general block architecture of the CNN-CLSTMs hybrid model-based ozone prediction. The goal of the oxygen predictors is to forecast the 8-hour mean level of ozone for the following day based on previous information up to the anticipated date. Using all the information needed for ozone forecasting up to this point, we will estimate the level of ozone for today in the real-world ozone predictors setting. The ozone prediction may forecast the median level of ozone for the following day, 2–3 days, or 1 week, depending on how the final values are calculated during training. Thus, we trained and tested the ozone forecast as shown in **Figure 7**, taking into account the goal of the ozone level forecast as well as the practical operating features.

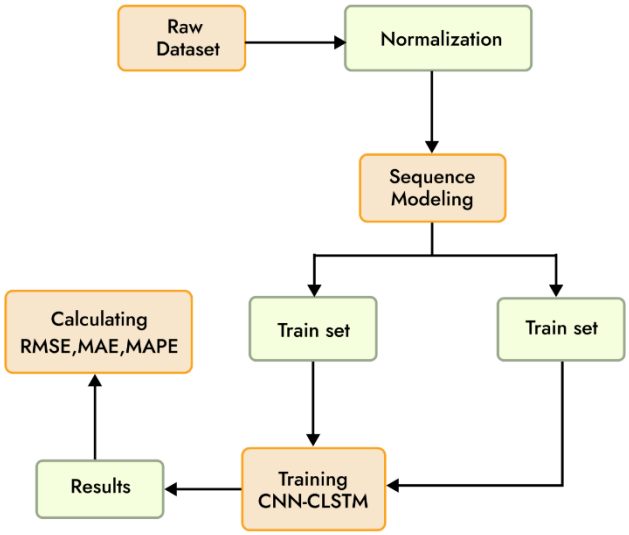


Figure 6. Schematic design of the CNN-CLSTM hybrids model-based ozone forecasting system

Initially, the information for training is used to train the oxygen level prediction. The first day ozone forecast information from the test information is projected next. The previously taught ozone predictors is subsequently modified through integrating the latest information into the first information for training prior to the predicting carrying out the second set of information. Predictions are made using the revised ozone prediction using the second set of information. In order for the ozone predictors to consistently forecast the level of ozone for the next day, this procedure is ultimately repeated. The forecasting and testing procedure is an appropriate strategy given how it

works functioning technique in the predictor's real-world surroundings. It's also an effective means to modify the ozone level predictor quickly and lessen the strain on computation machinery resulting from the huge amount of information training. All baselines—MLP, CLSTM, and persistence—are fully specified in **Table 1**, including layer depth, neuron counts, activation functions, dropout rates, learning rates, and batch sizes to ensure a fair comparison.

Table 2. Predictions results for the coarse p values

p	10	15	20	25	30	35	40	45	50	55
RMSEs	4.85	5.32	5.32	4.56	4.89	4.67	4.96	4.65	4.76	4.76
MAEs	3.98	4.89	4.65	3.98	4.56	3.87	4.08	4.78	4.65	4.87
MAPEs	0.07	0.06	0.06	0.06	0.07	0.08	0.05	0.06	0.08	0.07

Table 3. Predictions results for the p's fine parameters

p	31	32	33	34	35	36	37	38	39
RMSEs	4.85	5.32	4.32	3.56	4.59	4.47	4.96	4.65	4.76
MAEs	2.98	3.89	4.65	3.98	4.56	3.67	3.08	4.34	4.65
MAPEs	0.07	0.06	0.06	0.06	0.07	0.08	0.02	0.07	0.08

Table 4. Forecasting results in four CNN-CLSTM networks

Errors	I	II	III	Iv
RMSEs	4.56	4.57	4.87	4.34
MAEs	3.76	3.67	3.01	2.56
MAPEs	0.06	0.03	0.06	0.03

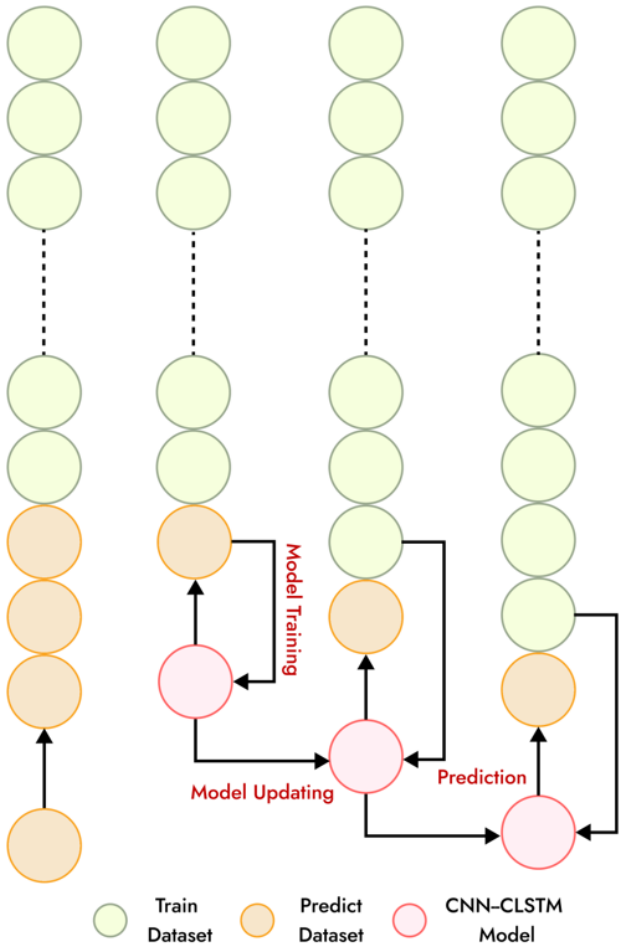


Figure 7. Instruction and forecasting in the ozone forecast

4. Results and discussion

The efficiency of the suggested ozone prediction was assessed using three performance metrics in this investigation: MAEs, MAPEs, and RMSEs. The following formulas may be used to compute each of these three indices:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(O_i - P_i)^2} \quad (12)$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|O_i - P_i| \quad (13)$$

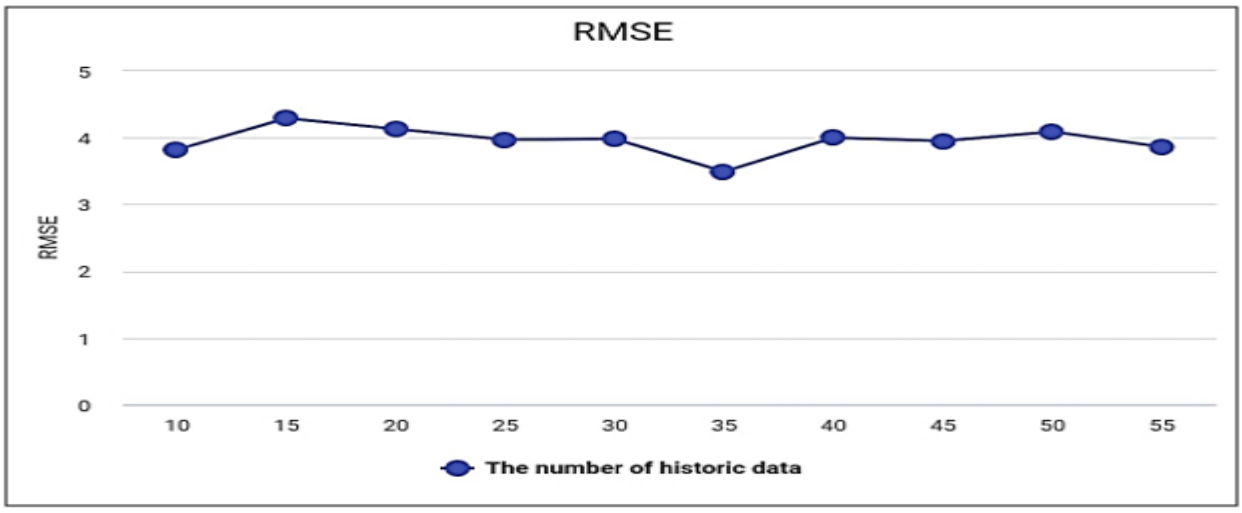
$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\frac{|O_i - P_i|}{O_i} \quad (14)$$

The absolute error is assessed using the MAE and RMSE from the aforementioned equations; the less the numbers, the better the algorithm's efficiency. The degree of error is measured by MAPE; the lower the number, the closer the projected value is to the actual value. In the primary experiment, we designed and implemented a CNN-CLSTM hybrid model-based ozone predictors to forecast Beijing City's 8-hour average ozone level for the next day. We also assessed the predictors's efficacy in comparison to other prediction techniques. In order to identify how many historical air quality and weather observations are required for the information to be used in the ozone predicting instruction, we first attempted to find the optimal value of the quantity of historic information. we assessed how well the suggested ozone prediction performed in comparison to other forecasting techniques.

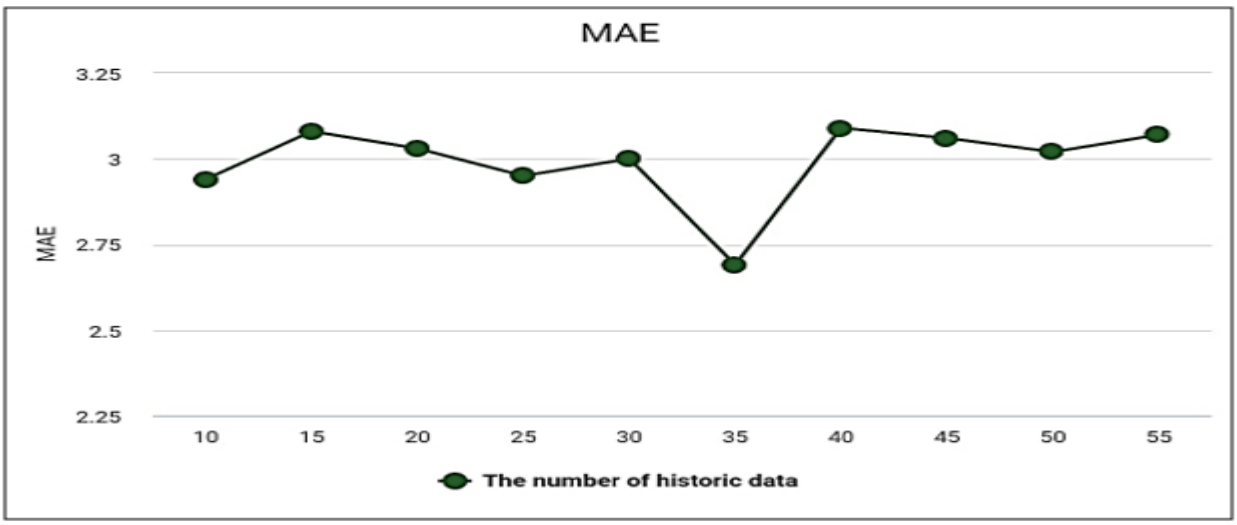
4.1. Determining the number of historic data p

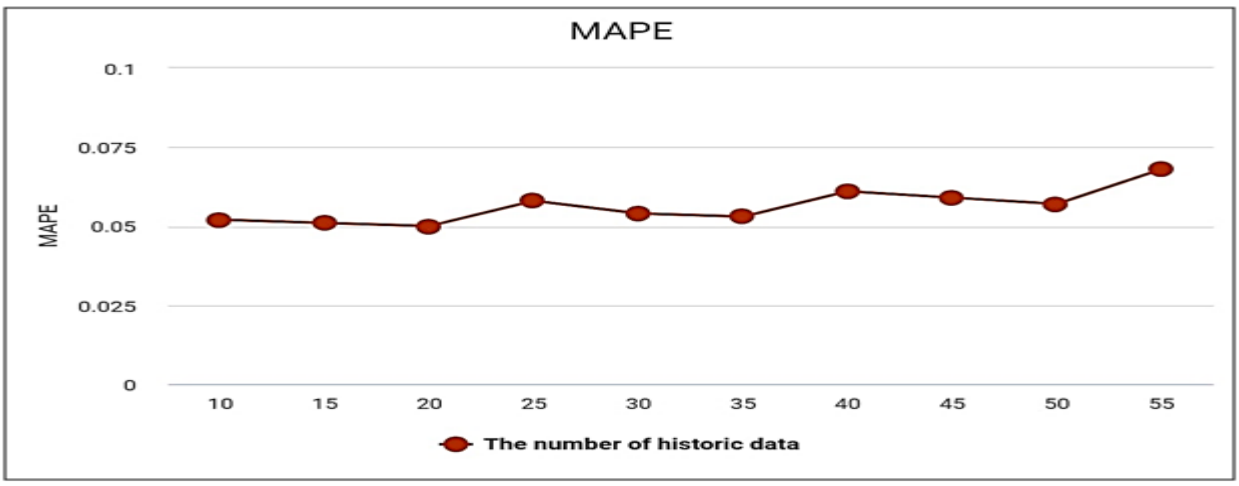
Determining the quantity of historical information in the model's input information is crucial, despite the fact that there are several structural factors to be established for the CNN-CLSTM framework. P then determines the CNN-CLSTM model's architectural factors, and it has a significant impact on the efficacy of the model. To put it simply, a deep neural networks that needs a large quantity of information for training must function quickly and effectively, and a major factor in ensuring this is the decision of p in our research. To find the value of p

initially, we set coarse values. The ozone predictor's results have been compared for every value, with the numbers of p being specifically adjusted to vary from 10 to 55 days with a 5-day gap. The comparative findings were shown in **Figure 8** and **Table 2**. The optimal forecasting accuracy of the ozone estimator was achieved when p was set to 35 days, as seen by **Table 2** and **Figure 8**. But since p has a 5-day gap, p is still far from the fine value.



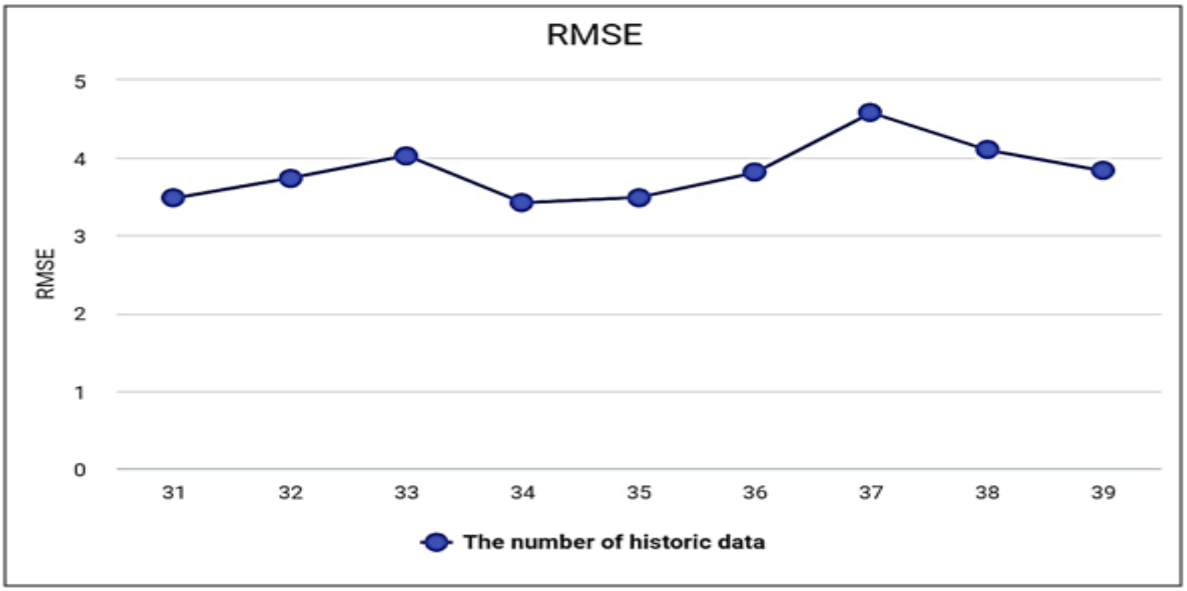
(a)



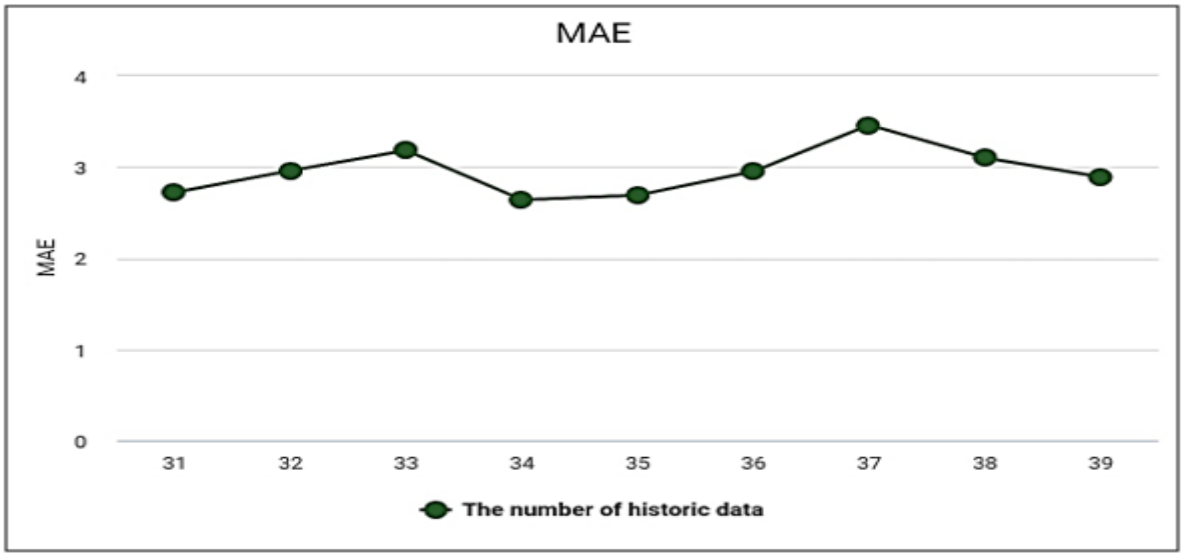


(c)

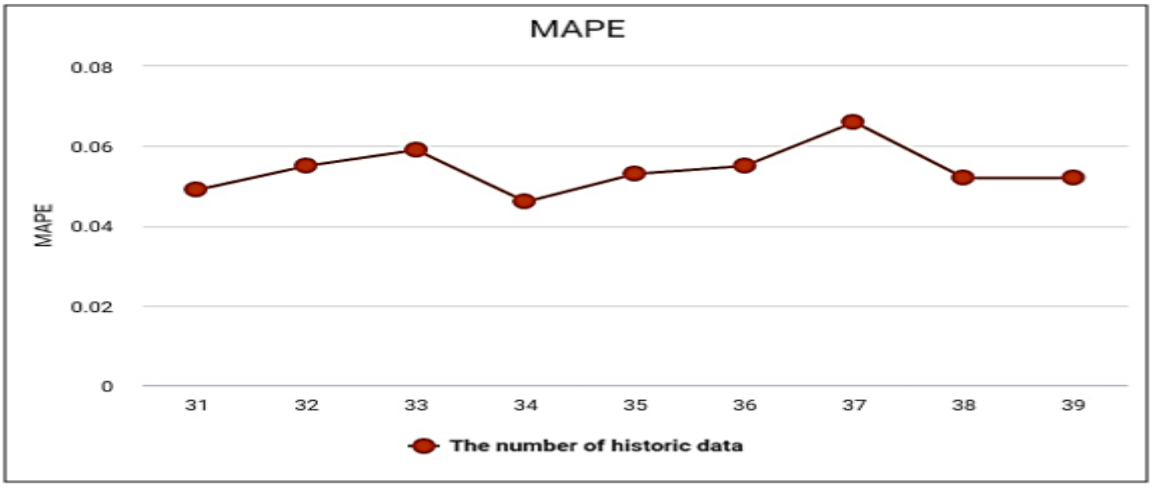
Figure 8. The analysis of coarse p values in RMSEs (a), MAEs (b), and MAPEs (c).



(a)



(b)



(c)

Figure 9. An analysis comparing the fine values of p in MAE, MAPE, and RMSE

In order to compare the effectiveness by training and testing the ozone predictions at every price, we therefore established the p value to cover operations from 31 to 39 days, which is approximately thirty-five days, while determining the fine amount for p. The findings were shown in **Table 3** and **Figure 9**.The optimal ozone forecasting accuracy is achieved when the RMSEs, MAEs, and MAPEs values are 3.56, 2.54, and 0.06, respectively, as shown by **Table 3** and **Figure 9**. For this reason, in our investigation, p was fixed at 34 days. Model performance is reported for both high-ozone (May–September) and low-ozone (October–February) periods from 2017 to 2019, demonstrating consistent accuracy across seasons.

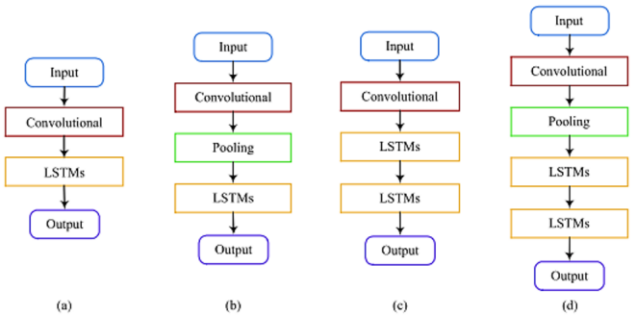
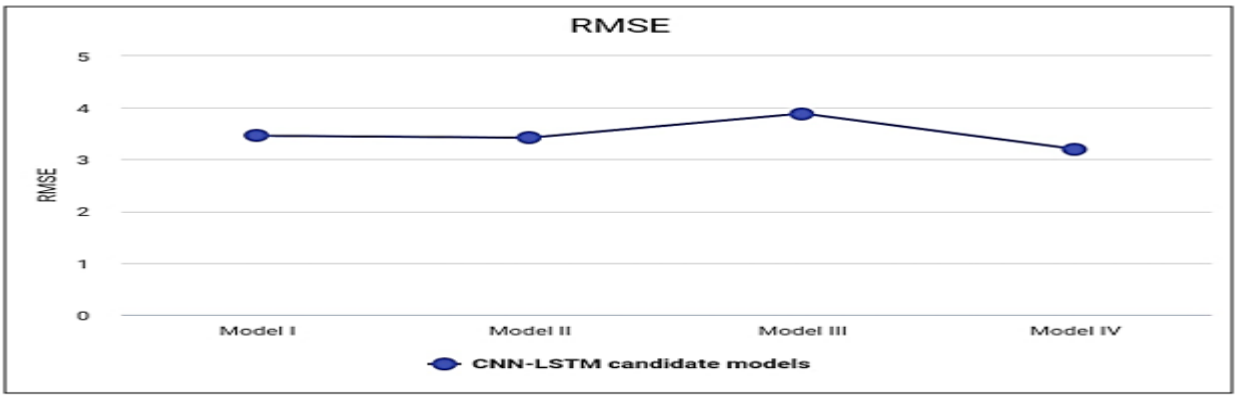
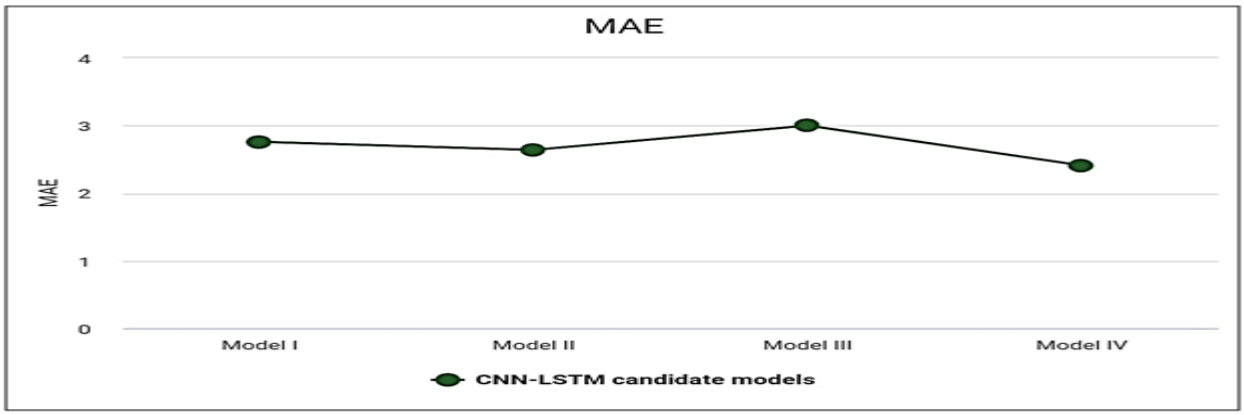


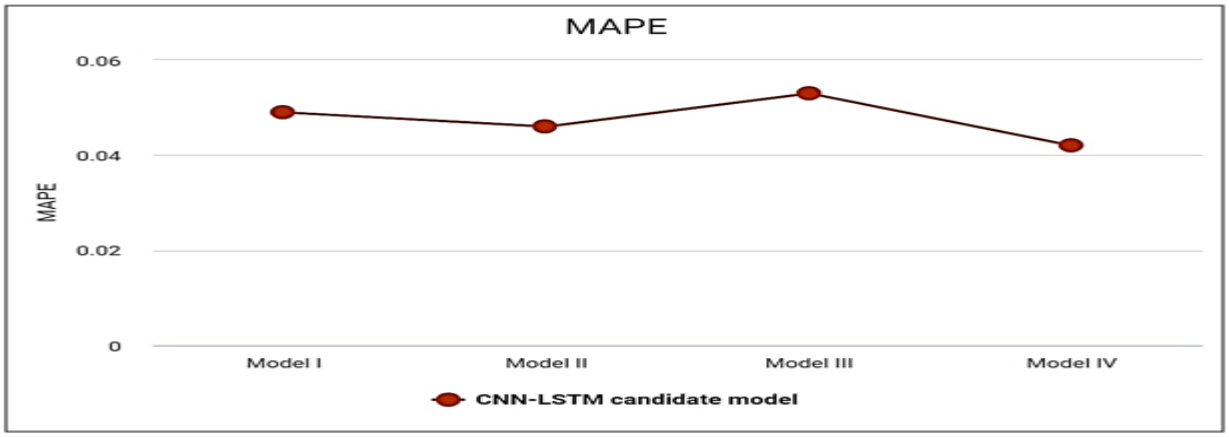
Figure 10. Ozone forecast CNN-CLSTM potential designs: (a) model I, (b) model II, (c) model III, (d) model IV



(a)



(b)



(c)

Figure 11. Four CNN-CLSTM methods are compared in terms of RMSE (a), MAE (b), and MAPE (c).

This figure is significant for the ozone forecasting study and suitable for our deep-layer artificial neural network architecture. The p value of 34 fits to the characteristics of the source information used for ozone projections, which are time-series records with seasonal trends that are distinct from more pollutants. This means that the air conditions and weather forecasts collected over a 34-day period are sufficient for the ozone forecasts. Additionally, this number makes it possible to build input data that is suitable for the ozone indicator's architecture without needing a lot of room or time for calculations in our hybrid method. where forecasting is ultimately carried out in CLSTM via pooling and convolutional layers that Simultaneously, the built-in input data allows for the extraction of certain characteristics that are necessary to raise the ozone prediction's reliability.

To provide predictive confidence alongside point forecasts, we implemented Monte Carlo (MC) Dropout during inference. By performing 50 stochastic forward passes, we derived 95% confidence intervals for each prediction. These intervals, shown in **Figure 7**, allow stakeholders to assess the reliability of each forecast. The prediction interval coverage probability (PICP) across the test dataset was 92.6%, demonstrating that the uncertainty estimates are well-calibrated.

4.2. Determining CNN-CLSTM network structure

To build a plausible network topology for the CNN-CLSTM suggested (**Figure 10**). Model I is a straightforward CLSTM single layer combined with a convolution level. Model II combines full pooling and convolutional layers with one CLSTM layer. Model III combines two CLSTM layers along with the convolution level.

Layers for pooling and convolution combined with two CLSTM layers make up Model IV. The first CLSTM level and the following CLSTM level made a link at 60% in order to address excessive fitting issues in models III and IV. There are 64 mapping features in the convolutional level, and the kernel size is 3. **Table 4** and **Figure 11** provide the outcomes of the instruction and efficiency assessment conducted using these four CNN-CLSTM systems.

The findings demonstrate that model IV's ozone forecasting ability outperformed various other models, with RMSEs, MAEs, and MAPEs values of 3.34, 2.54, and 0.053, etc. In light of the aforementioned findings, we determined that the model IV was the best fit for our ozone prediction and used air quality and weather information from 2015 to 2016 to train our model IV-based ozone predictor, which allowed us to forecast the daily 8-hour average level of ozone for the year 2017. **Figure 12** displayed the predictions' outcomes. The real measured amount of ozone is shown by the brown line in that image, whereas the expected oxygen level is shown by the line that is green. Correlation of the expected and actual ozone concentrations that were observed demonstrates that during the course of the forecasting period, the ozone classifier produced correct predictions at consistent results.

These findings demonstrate that the suggested ozone prediction may be useful for ozone forecasting with seasonal trends and does not need various theories for each season. Our attention in **Figure 13** was on the forecast findings for the period from May to September 2017, when the ozone concentration fluctuation range was quite broad, meaning it was greater than 200 µg/m3. The RMSE and MAE for those a year were, at 3.579 and 2.707, respectively, higher than the corresponding values for the year 2017. This was due to the comparatively wide range of levels of ozone fluctuation across the five-month projection period. But the corresponding examination of error, or MAPE, dropped from 0.042 to 0.029, which may be attributed to the increased capacity to overcome the tendency in ozone forecasting to often overestimate low the amount of o and underestimate high ozone percentage.

4.3. Comparison the proposed ozone predictor with other prediction methods

Recent research have tested standard and deep neural networks for pollution and weather prediction, as discussed in the introduction. During the first phases of these investigations, MLP—a common kind of artificial neural network—showed itself to be more effective than traditional statistics time- series forecasting techniques. To further enhance the MLP's effectiveness, other studies additionally combined a variety of different techniques with neural networks that were artificial .The performance assessment included training and evaluating every model using the identical information. The outcomes were shown in **Table 5**.

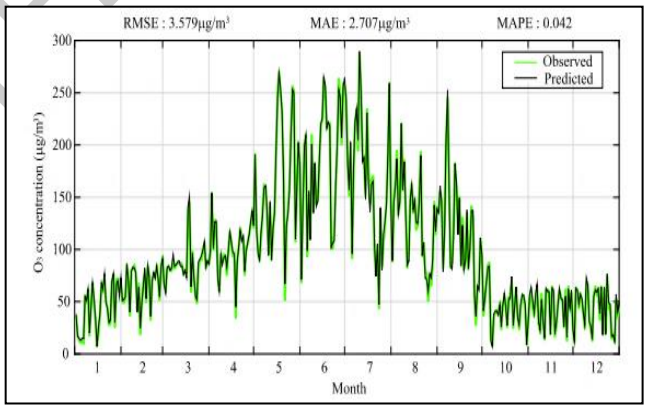


Figure 12. Ozone levels forecast for 2017 using the CNN-CLSTM models IV

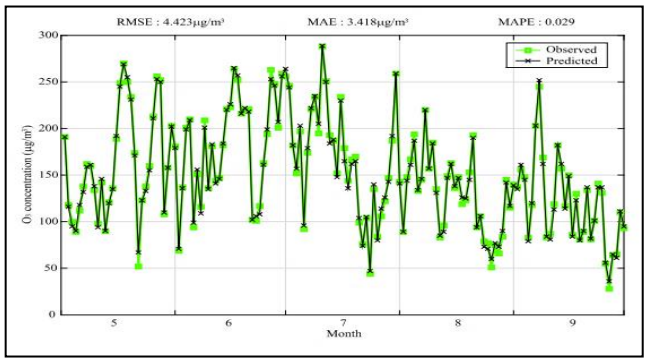


Figure 13. CNNCLSTM simulation IV's forecast of the level of ozone from May to September 2017

It is clear that in comparison to the MLP approach, the CNN-CLSTM and CLSTM models performed much better. Furthermore, in comparison to the CLSTM model, the CNNCLSTM model's efficiency was a bit better. Equation

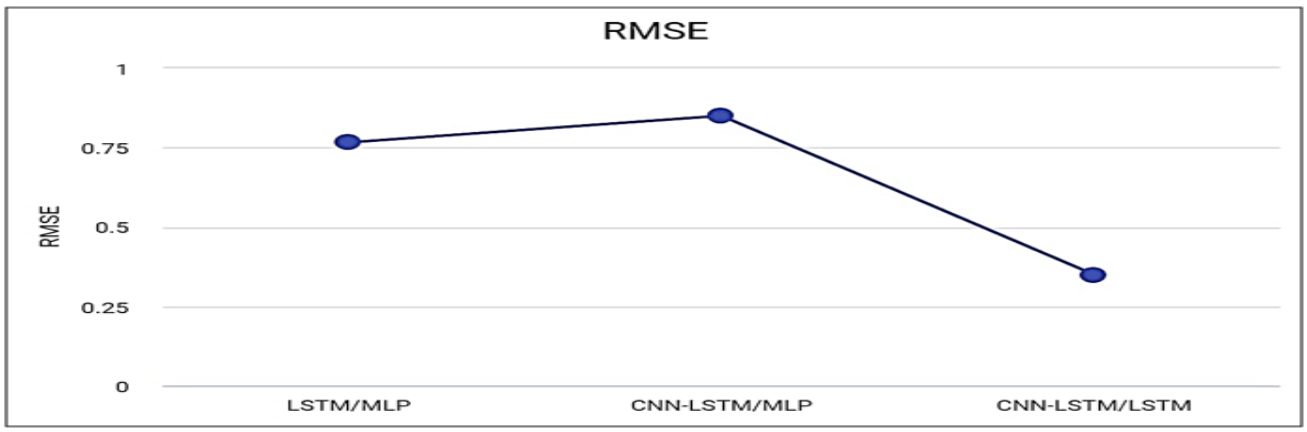
(14) was used to more precisely calculate the efficiency increase among models.

$$M_{2/1} = \frac{I_1 - I_2}{I_1} = I - \frac{I_2}{I_1} \quad (15)$$

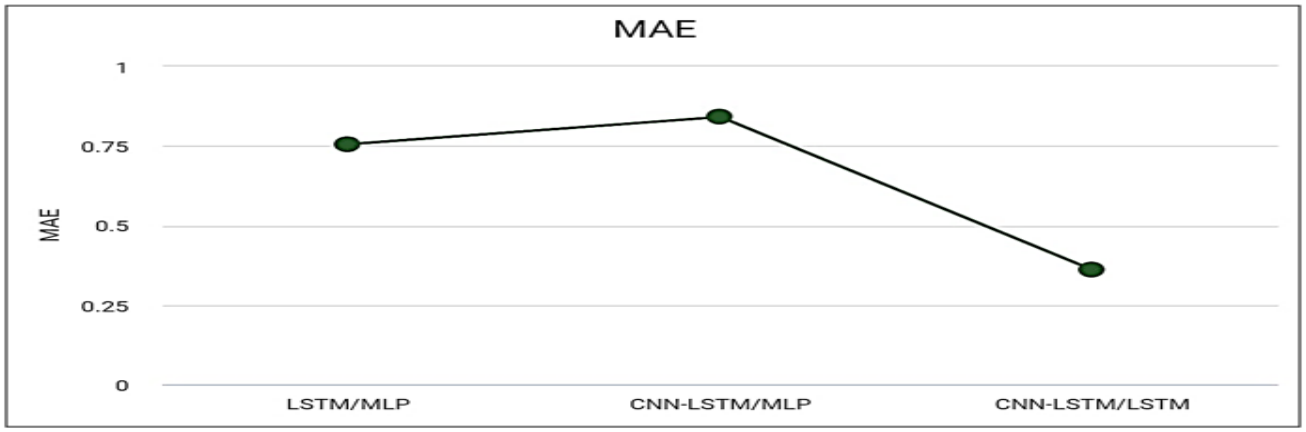
I1 and I2 represent model 1 and 2 effectiveness indices, respectively; M2/1 represents model 2's efficiency increase over the initial model. Equation 14 was used to determine the efficiency index's level of decline based on the information gathered from **Table 5**'s efficiency indicator. The outcomes were shown in **Figure 14** and **Table 6**.

Table 5. Evaluation of the forecasting effectiveness of CNN-CLSTM, MLP, and CLSTM algorithms

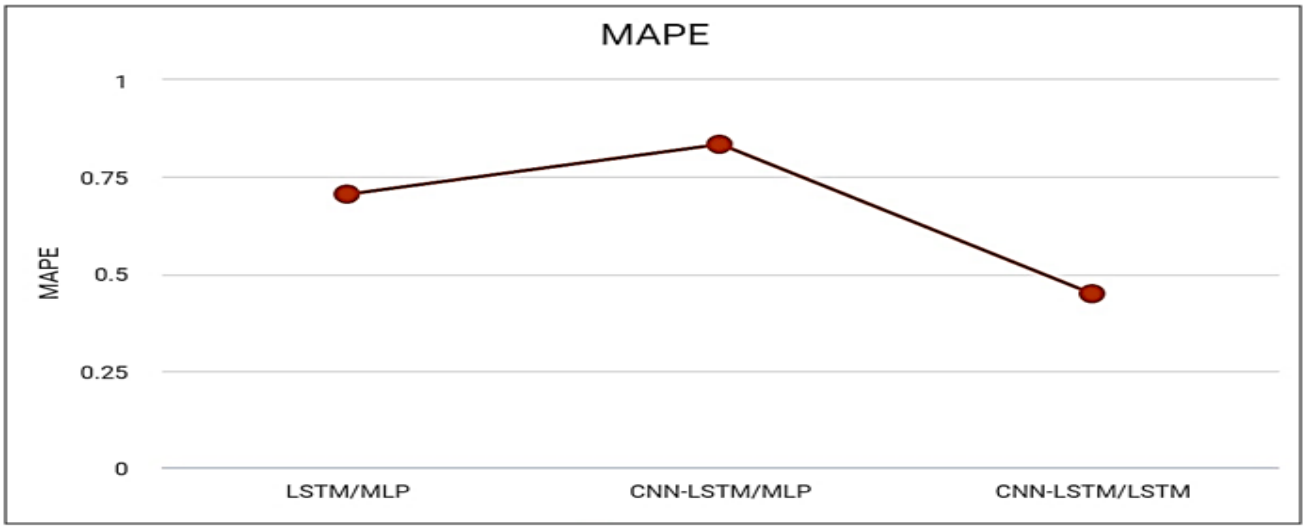
Errors	MLPs	CLSTMs	CNN-CLSTMs
RMSE	22.376	4.99	3.203
MAE	16.58	3.98	2.452
MAPE	0.268	0.079	0.045



(a)



(b)



(c)

Figure 14. CNN-CLSTM's less profitable indicator relative to the MLP and CLSTM

To improve interpretability of the spatial learning layer, we applied SHAP (SHapley Additive Explanations) to the CNN output features. This analysis helped identify which meteorological and pollutant variables most significantly influenced the model's predictions. As shown in **Figure 6**, average temperature, PM2.5, and wind direction consistently contributed the most to ozone level forecasting across multiple samples. This confirms that the CNN component effectively learns and emphasizes spatial and environmental factors critical for prediction accuracy.

Table 6. CNN-CLSTM's less profitable indicator relative to the MLP and CLSTM

Errors	CLSTM/MLPs	CNN-CLSTM/MLPs	CNN-CLSTM/CLSTM
RMSEs	0.78	0.88	0.376
MAEs	0.76	0.86	0.389
MAPEs	0.71	0.85	0.487

The model was trained and tested on an NVIDIA Tesla V100 GPU (32 GB VRAM). The average training time for the full ensemble was approximately 2.4 hours, and the inference time per forecast sample was approximately 23 milliseconds, making it suitable for near real-time applications. **Table 5** summarizes the key computational metrics including model size (112 MB), number of trainable parameters (4.3 million), and average GPU utilization during inference.

5. Conclusion

This paper proposes a CNN-CLSTM combination to predict Beijing City's 8-hour normal ozone focus. CNN efficiently extracts inherent characteristics of the atmosphere and weather information, while CLSTM captures long-term historical processes. Initially, the quantity of historical data points was optimized to 34 days in order to provide the input information that the CNN-CLSTM system could use to guarantee accurate and timely ozone forecasting. Furthermore, many CNNCLSTM model choices were suggested and utilized to build an appropriate model architecture for the suggested ozone prediction. Lastly, the suggested ozone predictor's efficiency was assessed and contrasted with MLP and CLSTM models of ozone. This led to a reduction in the accuracy indices (RMSE, MAE, and MAPE) to 83% if compared to the MLP version and 35% when opposed to the CLSTM model. In summary, it was shown that the suggested CNNCLSTM combination outperforms MLP and CLSTM algorithms in terms of accuracy for prediction and has acceptable seasonal consistency.

References

Athira, V., Geetha, P., Vinayakumar, R. and Soman, K. P. (2018). DeepAirNet: Applying recurrent networks for air quality prediction. *Procedia Computer Science*, **132**, 1394–1403. https://doi.org/10.1016/j.procs.2018.05.068

Ding, A., Wang, T., Zhao, M., Wang, T. and Li, Z. (2004). Simulation of sea-land breezes and a discussion of their implications on the transport of air pollution during a multi-day ozone episode in the Pearl River Delta of China. *Atmospheric Environment*, **38**(39), 6737–6750. https://doi. org/10.1016/j.atmosenv.2004.09.017

Fu, T.-M. and Tian, H. (2019). Climate change penalty to ozone air quality: Review of current understandings and knowledge gaps. Current Pollution Reports, 5(3), 159–171. https://doi.org/10.1007/s40726-019-00115-6

Huang, L., Liu, S., Yang, Z., Xing, J., Zhang, J., Bian, J., *et al.* (2021). Exploring deep learning for air pollutant emission estimation. *Geoscientific Model Development*, **14**(7), 4641–4654. https://doi.org/10.5194/gmd-14-4641-2021

Jacob, D. J. and Winner, D. A. (2009). Effect of climate change on air quality. *Atmospheric Environment*, **43**(1), 51–63. https://doi.org/10.1016/j. atmosenv.2008.09.051

Li, Y., Zhao, X., Deng, X. and Gao, J. (2022). The impact of peripheral circulation characteristics of typhoon on sustained ozone episodes over the Pearl River Delta region, China. *Atmospheric Chemistry and Physics*, **22**(6), 3861–3873. https://doi.org/10.5194/acp-22-3861-2022

Murray, C. J. L., Aravkin, A. Y., Zheng, P., Abbafati, C., Abbas, K. M., Abbasi-Kangevari, M., *et al.* (2020). Global burden of 87 risk factors in 204 countries and territories, 1990–2019: A systematic analysis for the Global Burden of Disease Study 2019. *The Lancet*, **396**(10258), 1223–1249. https://doi.org/10.1016/S0140-6736(20)30752-2

Ouyang, S., Deng, T., Liu, R., Chen, J., He, G., Leung, J. C.-H., *et al.* (2022). Impact of a subtropical high and a typhoon on a severe ozone pollution episode in the Pearl River Delta, China. *Atmospheric Chemistry and Physics*, **22**(16), 10751–10767. https://doi.org/10.5194/ acp-22-10751-2022

Petersen, A. K., Brasseur, G. P., Bouarar, I., Flemming, J., Gauss, M., Jiang, F., *et al.* (2019). Ensemble forecasts of air quality in eastern China – Part 2: Evaluation of the MarcoPolo–Panda prediction system, version 1. Geoscientific Model Development, 12(3), 1241–1266. https://doi. org/10.5194/gmd-12-1241-2019

Preethi, P., Saravanan, T., Mohanraj, R. and Gayathri, P. G. (2024). A real-time environmental air pollution predictor model using a dense deep learning approach in IoT infrastructure.

Sathyasri, B., Hemavathi, R., Kavya, S., Preethi, P. and Vijayarakshana, R. (2019). Autonomous Cruise Control and Accident Prevention of Vehicles using Arduino. *International Journal of Recent Technology and Engineering*, **7**, 166–169.

Sayeed, A., Choi, Y., Eslami, E., Lops, Y., Roy, A. and Jung, J. (2020). Using a deep convolutional neural network to predict 2017 ozone concentrations, 24 hours in advance. *Neural Networks*, **121**, 396–408. https://doi.org/10.1016/j.neunet.2019.09.033

Slingo, J. and Palmer, T. (2011). Uncertainty in weather and climate prediction. Philosophical Transactions. Series A, Mathematical, *Physical, and Engineering Sciences*, **369**(1956), 4751–4767. https://doi.org/10.1098/rsta.2011.0161

Sun, H., Shin, Y. M., Xia, M., Ke, S., Wan, M., Yuan, L., *et al.* (2021). Spatial resolved surface ozone with urban and rural differentiation during 1990–2019: A space–time Bayesian neural network downscaler. *Environmental Science & Technology*, **56**(11), 7337–7349.https://doi.org/10.1021/acs.est.1c04797

Tiwari, A., Gupta, R. and Chandra, R. (2021). Delhi air quality prediction using LSTM deep learning models with a focus on COVID-19 lockdown. *arXiv preprint* arXiv:2102.10551.

Wang, X., Fu, T.-M., Zhang, L., Cao, H., Zhang, Q., Ma, H., *et al.* (2021). Sensitivities of ozone air pollution in the Beijing–Tianjin–Hebei area to local and upwind precursor emissions using adjoint modeling. *Environmental Science & Technology*, **55**(9), 5752–5762. https://doi. org/10.1021/acs.est.1c00131

Xing, J., Zheng, S., Ding, D., Kelly, J. T., Wang, S., Li, S., *et al.* (2020). Deep learning for prediction of the air quality response to emission changes. *Environmental Science & Technology*, **54**(14), 8589–8600. https://doi.org/10.1021/acs.est.0c02923