International Journal of Engineering and Information Systems (IJEAIS)
ISSN: 2643-640X
Vol. 4, Issue 8, August – 2020, Pages: 88-101

An Automated LSTM-based Air Pollutant Concentration Estimation of Dhaka City, Bangladesh

Rehnuma Karim[1] and Taki Hasan Rafi[2]

[1]Department of Environmental Science, Jahangirnagar University, Savar, Bangladesh
Email: rehnumakarimrinky@gmail.com
[2]Department of Electrical and Electronic Engineering, Ahsanullah University of Science and Technology, Dhaka, Bangladesh
Email: takihasanrafi@gmail.com

Abstract**—***Dhaka is one of the mega-cities around the global. It has around 20 million population. Because of the outrageous population in less area, Dhaka is considered as one of the air polluted urban communities. So in this regard, air pollutant projection determining is a viable technique for ensuring general health by giving an early admonition against destructive air pollutants. It is a challenging task for researchers. However, existing automated strategies for air pollutant projection forecast is depend on term data acknowledgement. Which occurs forecast errors in many cases. In this paper, an efficient time-series artificial neural network model, Long-short term memory (LSTM) has been utilized for air pollutant concentration forecasting. Monthly* $PM_{2.5}$*,* PM_{10}*,* SO_2*,* NO_2*,* CO_2 *and* O_3 *concentration data collected at 3 air quality monitoring stations in Dhaka city from Jan-2013 to Jun-2020 were utilized to approve the viability of the proposed LSTM model. Additionally, the proposed model has been compared by autoregressive moving average (ARMA) and time-delay neural network (TDNN) model. To evaluate the viability of proposed model, mean absolute percentage error (MAPE) method has been utilized.*

Keywords—Air pollution, LSTM, Dhaka city; $PM_{2.5}$, PM_{10}.

1. INTRODUCTION

The earth consists of different spheres, and the atmosphere is one of the most important spheres of them all Expanded ignition of petroleum derivatives in the only remaining century is answerable for the dynamic change in the atmospheric creation. Air toxins, for example, carbon monoxide (CO), sulfur dioxide (SO_2), nitrogen oxides (NO_2), unstable natural mixes (VOCs), ozone (O_3), overwhelming metals, and respirable particulate issue ($PM_{2.5}$ and PM_{10}), contrast in their compound arrangement, response properties, outflow, time of deterioration and capacity to diffuse in long or short separations [6]. Presently the grouping of sulfur dioxide has diminished strikingly, consideration has moved to ozone, nitrogen dioxide, and particulates [5].Worldwide air-quality issues exist just in regard to those poisons whose barometrical lifetimes are sufficiently long (on the request for 1 week) for them to be moved at any rate to another landmass [3].One of our time's greatest sources is air pollution, not just for its impact on climate change but also for its impact on public and individual health by increasing morbidity and mortality [1]. These changes in climate and air quality have a quantifiable impact, not only on the morbidity but also on the mortality for disease Global earth temperature has markedly risen over the last 5 decades thanks to the rise in gas emissions. Warming from anthropogenic-derived greenhouse gas has consequences, including temperature change and public health risks. Every year millions of people get killed by air pollution. The World Health Organization estimates that particulate (PM) pollution contributes to approximately 800,000 premature deaths annually, ranking it the 13th leading reason for mortality worldwide [4].As indicated by air contamination program of the World Health Association (WHO), 91% of the total populace breathes in polluted air, and about 4.2 million passing happen each year on account of surrounding air contamination. It uncovers that air contamination is causing more than 33% passing from lung malignant growth, strokes, and ceaseless respiratory ramifications [8].

Air pollution has two types of effect on the human body, acute and chronic. It varies from minor upper respiratory irritation to chronic respiratory and heart disease, carcinoma, acute respiratory infections in children and bronchitis in adults, aggravating pre-existing heart and lung disease, or asthmatic attacks [6]. Air contamination is likewise connected with expanded irritation and tension [7].With the expansion of sources and kinds of air contaminants, the multifaceted nature of toxin fixation expectation has expanded [2].

Air pollution can be categorized into two different terms, indoor and outdoor air pollution. There are numerous wellsprings of indoor air contamination. Tobacco smoke, cooking and warming machines, and fumes from building materials, paints, furniture, and so forth cause contamination inside structures. Outdoor air is regularly alluded to as encompassing air. The basic wellsprings of open-air contamination are discharges brought about by ignition forms from engine vehicles, strong fuel copying and industry. The most normal air toxins of encompassing air include Particulate matter (PM_{10} and $PM_{2.5}$). Outdoor air contamination costs a large number of unexpected losses yearly, for the most part, because of the anthropogenic fine particulate issue (or $PM_{2.5}$) [11].

Fine PM is arranged into PM_{10} and $PM_{2.5}$ dependent on molecule distance across, where PM_{10} and $PM_{2.5}$are particles with measurements <10 and <2.5 m, individually. PM incorporates dust, pollen, sediment, smoke, what's more, fluid beads that hurt the respiratory framework, causing respiratory side effects including sporadic pulse, hacking, aviation route aggravation, strange lung work, breathing difficulty, heart assault, stroke-related illnesses, and asthma [10]. Recent studies recommend that relative increments in day by day mortality per 10-µg/m3 increment in PM10 are commonly comparable among North American and Western European locales and a couple of creating nations where studies have been attempted [9].

Air pollution is a major issue in most of the developing countries, Bangladesh is one of them. The growing economic development and technological revolution is the key reason behind the most polluted air in Dhaka city, the capital of Bangladesh. Particulate matter focuses in Dhaka have been estimated constantly since December 1996 [13]. Dhaka has a population of around 22 million (in roughly 1700 km zone) is probably the greatest city of the creating nations. It is growing quickly because of urbanization and high convergence of individuals from rustic zones. Emanations from different curious sorts of diesel traffic vehicles, for instance, three-wheeled auto-rickshaws, and badly maintained autos, contribute most to air contamination issues [12].

2. VARIOUS AIR POLLUTANTS

2.1 CO_2

2.1.1 Sources

Procedures or districts that predominately produce barometrical carbon dioxide are called sources. Carbon dioxide is added to the environment normally when creatures breathe or break down, carbonate rocks are endured, woods fires happen, and volcanoes emit. Carbon dioxide is likewise added to the climate through human exercises, for example, the consuming of non-renewable energy sources and backwoods and the creation of concrete.

Carbon dioxide is added to the climate by human exercises. At the point when hydrocarbon energizes (for example wood, coal, flammable gas, gas, and oil) are scorched, carbon dioxide is discharged. During ignition or consuming, carbon from non-renewable energy sources consolidates with oxygen noticeable all around to shape carbon dioxide and water fume. These common hydrocarbon energizes originate from once-living creatures and are produced using carbon and hydrogen, which discharge carbon dioxide and water when they copy. Fig. 1, Fig. 2 and Fig. 3 show the various sources of CO_2.

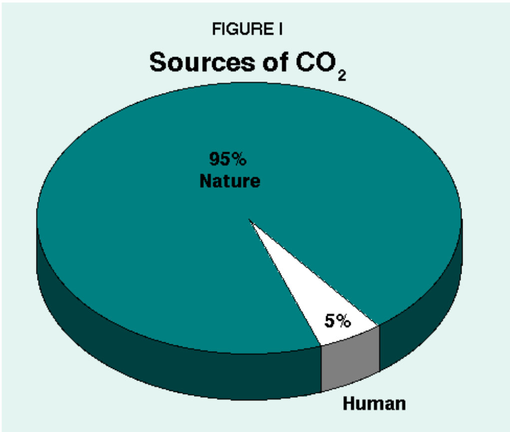

Fig. 1: Sources of CO_2

2.1.2 Effect on Environment

An Earth-wide temperature boost has been progressively connected with the commitment of CO_2. At present, it is assessed that CO_2 contributes about half to the anthropogenic nursery effect. The danger of environmental change because of emanations of CO_2 from petroleum products is thought of to be the principal ecological danger from the current vitality framework. Other ecological issues are fermentation, and scattering of metals beginning from non-renewable energy sources [14].

ISSN: 2643-640X
Vol. 4, Issue 8, August – 2020, Pages: 88-101

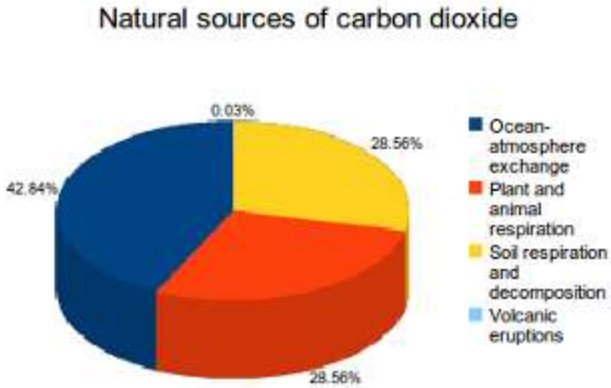

Fig. 2: Natural sources of CO_2

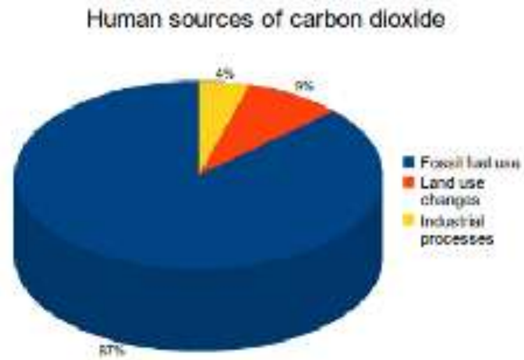

Fig. 3: Human sources of CO_2

2.1.3 Effect on Human Health

Table 1 shows the details of CO_2 effects on human health [15].

CO_2	*Effects of CO_2 on Human Health*
1%	*Respiratory rate upto 37%*
1.6%	*Respiratory rate upto 46%*
2%	*Respiratory rate upto 50%; High brain blood flow*
3%	*Exercise tolerance low*
5%	*Respiratory rate upto 100%; Dizziness*
7.2%	*Respiratory rate upto 200%; Dizziness*
8-10%	*Severe dizziness; sweating*
10% (Above)	*Unbearable dyspnea followed by vomiting, hypertension etc*

2.1.4 Present Situations, Acceptance Limits, Future Aspects

The worldwide normal air carbon dioxide in 2018 was 407.4 parts for each million (ppm for short), with a scope of vulnerability of give or take 0.1 ppm. Carbon dioxide levels today are higher than anytime in at any rate the previous 800,000 years. In actuality, the last time the climatic CO_2 sums were this high was in excess of 3 million years back, when temperature was 2°–3°C (3.6°–5.4°F) higher than during the pre-modern period, and ocean level was 15–25 meters (50–80 feet) higher than today. The yearly pace of increment in barometrical carbon dioxide in the course of recent years is around multiple times quicker than past normal increments, for example, those that happened toward the finish of the last ice age 11,000-17,000 years prior.

The protected degree of carbon dioxide in the environment is 350 ppm. The best way to arrive is to promptly progress the worldwide economy away from petroleum products and into sustainable power source, vitality effectiveness, and supportable cultivating rehearses. 400 ppm – 350 ppm.

On the off chance that ebb and flow ozone-depleting substance (GHG) concentrations stay steady, the world would be focused on a few centuries of increasing global mean temperatures and ocean level ascent (1–3). By differentiation, close end of anthropogenic CO_2 outflows would be required to create diminishing GHG fixations predictable with adjustment of mean temperatures (4–6) [15].

2.2 O_3

Ground-level or "terrible" ozone isn't radiated straightforwardly into the air, however is made by synthetic responses between oxides of nitrogen (NO_2) and unpredictable natural mixes (VOC) within the sight of daylight. Outflows from modern offices and electric utilities, engine vehicle fumes, fuel fumes, and concoction solvents are a portion of the significant wellsprings of NO_2 and VOC.

2.2.1 Sources

Photochemical oxidants are shaped when daylight falls on a blend of synthetic substances noticeable all around. Ozone is one of the primary photochemical oxidants. Different synthetic compounds, for example, formaldehyde are additionally found and, similar to ozone. Ozone is a gas that is formed when nitrogen oxides respond with a gathering of air contaminations known as 'receptive natural substances' within the sight of daylight. The synthetic compounds that respond to frame ozone originate from sources, for example, engine vehicle fumes, oil refining, printing, petrochemicals, grass cutting, avionics, bushfires and consuming off. Engine vehicle fumes exhaust produce as much as 70% of the nitrogen oxides and half of the natural synthetic concoctions that structure ozone.

2.2.2 Effects on Environment

Ground-level ozone can easily affect plants and environments. These impacts include:

- Meddling with the capacity of touchy plants to create and store food, making them progressively powerless to specific infections, creepy crawlies, different toxins, rivalry and unforgiving climate.
- Harming the leaves of trees and different plants, contrarily affecting the presence of urban vegetation, just as vegetation in national parks and diversion regions and
- Decreasing woods development and harvest yields, conceivably affecting species assorted variety in biological systems.

2.2.3 Effects on Human Health

Breathing ozone can trigger an assortment of medical issues including chest torment, hacking, throat aggravation, and blockage. It can intensify bronchitis, emphysema, and asthma. Ground-level ozone likewise can diminish lung work and arouse the linings of the lungs. Rehashed presentation may for all time scar lung tissue.

2.2.4 Present Situations, Acceptance Limits, Future Aspects

Normal common foundation concentrations of ground-level ozone are around 30–100 micrograms for each cubic meter (µg/m3). Present moment (60 minutes) mean surrounding focuses in urban regions may surpass 300–800 µg/m3.

The regular measure of ozone in the lower climate is by and large around 0.04 parts per million (ppm), and that sum isn't unsafe to human wellbeing. On the off chance that no edge is accepted, the expansion in evaluated yearly O_3 fixations somewhere in the range of 2003 and 2020 outcomes in a 15% (11272 to 12930) increment in passing which tumbles to 8% (11272 to 12140) if atmosphere is held at 2003 conditions. In the event that a limit of 35 ppb is accepted, the expansion is 51% (1582 to 2391), which tumbles to 14% (1582 to 1802) when atmosphere is held at 2003 conditions.(https://royalsociety.org/~/media/Royal_Society_Content/policy/publications/2008/7925.pdf, page 85)

2.3 NO_2

2.3.1 Sources

Nitrogen dioxide is not generally discharged legitimately into the air. Nitrogen dioxide structures when nitrogen oxide (NO) and other nitrogen oxides (NO_2) respond with different synthetics noticeable all around to frame nitrogen dioxide. The fundamental wellspring of nitrogen dioxide coming about because of human exercises is the ignition of petroleum products (coal, gas and oil) particularly fuel utilized in vehicles. It is likewise created from making nitric corrosive, welding and utilizing explosives, refining of petroleum and metals, business assembling, and food producing. Regular wellsprings of other nitrogen oxides incorporate volcanoes and microbes.

2.3.2 Effects on Environment

NO_2 is a gas, whose nursery sway for each atom premise is multiple times higher than that of carbon dioxide. When NO_2 atom spreads into the environment, it remains there for more than hundred years before it is destroyed normally. Nitrous oxide (NO_2) convergence of 300 ppb can prompts the warming of the Earth by 0.1 Watts per square meter. NO_2 responds right away with basic natural synthetic concoctions and even ozone, to shape different harmful items for example, nitrate radical and nitrosamines. Nitrogen oxides and sulfur dioxide respond with different components in the air results in corrosive downpours [16].

2.3.3 Effects on Human Health

The primary impact of taking in raised degrees of nitrogen dioxide is the improved probability of respiratory issues. Nitrogen dioxide excites the covering of the lungs, and it can lessen invulnerability to lung diseases. This can cause issues, for example, wheezing, hacking, colds, influenza and bronchitis. Expanded degrees of nitrogen dioxide can affect individuals with asthma since it can cause progressively continuous and increasingly serious assaults. Youngsters with asthma and more established individuals with coronary illness are most in danger.

2.3.4 Present Situations, Acceptance Limits, Future Aspects

On January 22, 2010, EPA strengthened the health- based National Ambient Air Quality Standard (NAAQS) for nitrogen dioxide (NO2). EPA set a 1-hour NO2 standard at the level of 100 parts per billion (ppb). EPA also retained the annual average NO2 standard of 53 ppb.

The current ACGIH recommendation is for a 3 ppm TWA and a 5 ppm STEL. The NIOSH REL is 1 ppm as a 15-minute short-term limit. **OSHA's** former PEL was 5 ppm as a ceiling value. The Agency proposed, and the final rule establishes, a permissible exposure limit for nitrogen dioxide of 1 ppm as a 15-minute STEL.

The advancement of ozone and nitrogen dioxide over Europe between the current day and a future period with a +2 °C the earth-wide temperature boost comparative with the pre-mechanical atmosphere was considered utilizing four disconnected science transport models, each determined by an alternate atmosphere model [17].

2.4 Particulate Matter

2.4.1 Sources

These particles come in numerous sizes and shapes and can be composed of many various synthetics. Some are radiated legitimately from a source, for example, building locales, unpaved streets, fields, smokestacks or flames. Most particles structure in the climate because of complex responses of synthetic compounds, for example, sulfur dioxide and nitrogen oxides, which are contaminations transmitted from power plants, ventures and vehicles.

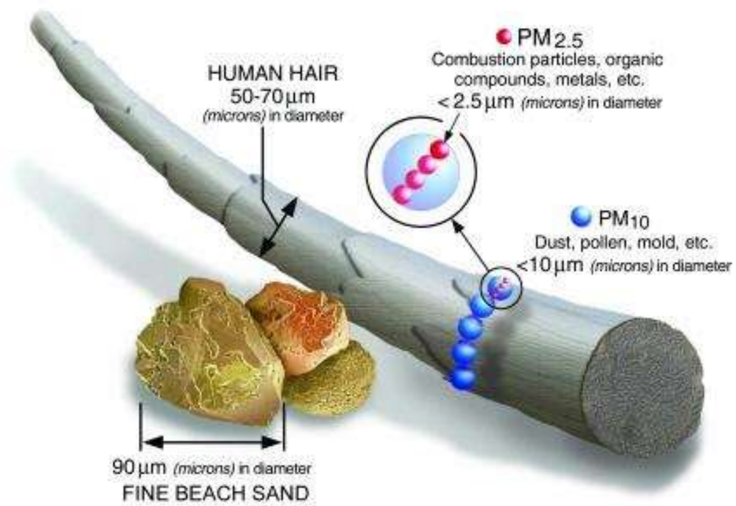

Fig. 4: Sources of particular matter

2.4.2 Effects on Environment

Particles can be persisted significant distances by wind and afterward choose ground or water. Contingent upon their concoction organization, the impacts of this settling may include:

- Making lakes and streams acidic.
- Changing the supplement balance in beach front waters and huge waterway bowls.
- Draining the supplements in soil.
- Harming touchy backwoods and homestead crops.
- Influencing the assorted variety of biological systems.
- Adding to corrosive downpour impacts.

2.4.3 Effects on Human Health

The size of particles is straightforwardly connected to their potential for messing wellbeing up. Little particles under 10 micrometers in breadth represent the best issues, since they can get profound into your lungs, and some may even get into your circulation system. Presentation to such particles can influence both your lungs and your heart. Various logical examinations have connected molecule contamination presentation to an assortment of issues, including:

- Unexpected passing in individuals with heart or lung malady
- Nonfatal cardiovascular failures
- Sporadic heartbeat
- Disturbed asthma
- Diminished lung work

- Expanded respiratory manifestations, for example, disturbance of the aviation routes, hacking or trouble relaxing.

Individuals with heart or lung illnesses, youngsters, and more seasoned grown-ups are well on the way to be influenced by molecule contamination presentation.

2.4.4 Acceptable Limits

Fine particulate matter ($PM_{2.5}$): 10 µg/m^3 annual mean and 25 µg/m^3 24-hour mean.
Coarse particulate matter (PM_{10}): 20 µg/m^3 annual mean and 50 µg/m^3 24-hour mean.

2.5 SO_2

2.5.1 Sources

SO_2 is a colorless gas with a sharp odour. It is produced from the burning of fossil fuels (coal and oil) and the smelting of mineral ores that contain sulfur. The main anthropogenic source of SO_2 is the burning of sulfur-containing fossil fuels for domestic heating, power generation and motor vehicles.

2.5.2 Effects of Environment, Human Health and Acceptable Limits

When SO_2 combines with water, it forms sulfuric acid; this is the main component of acid rain which is a cause of deforestation. SO_2 can affect the respiratory system and the functions of the lungs, and causes irritation of the eyes. Inflammation of the respiratory tract causes coughing, mucus secretion, aggravation of asthma and chronic bronchitis and makes people more prone to infections of the respiratory tract. Hospital admissions for cardiac disease and mortality increase on days with higher SO_2 levels. The acceptable limits are 20 µg/m^3 24-hour mean and 500 µg/m^3 10-minute mean.

3. DATA COLLECTION

Table 2: Air quality monitoring stations in Dhaka City.

City	*ID*	*Location*	*Latitude & Longitude*	*Monitoring Capacity*
Dhaka	*CAMS-1*	*Sangshad Bhaban Sher-e-Bangla Nagar*	*23.76N 90.39E*	*PM10, PM2.5, CO, SO2, NOX, O3*
	CAMS-2	*Firmgate*	*23.76N 90.39E*	*PM10, PM2.5, CO, SO2, NOX, O3*
	CAMS-3	*Darus-Salam*	*23.78N 90.36E*	*PM10, PM2.5, CO, SO2, NOX, O3*

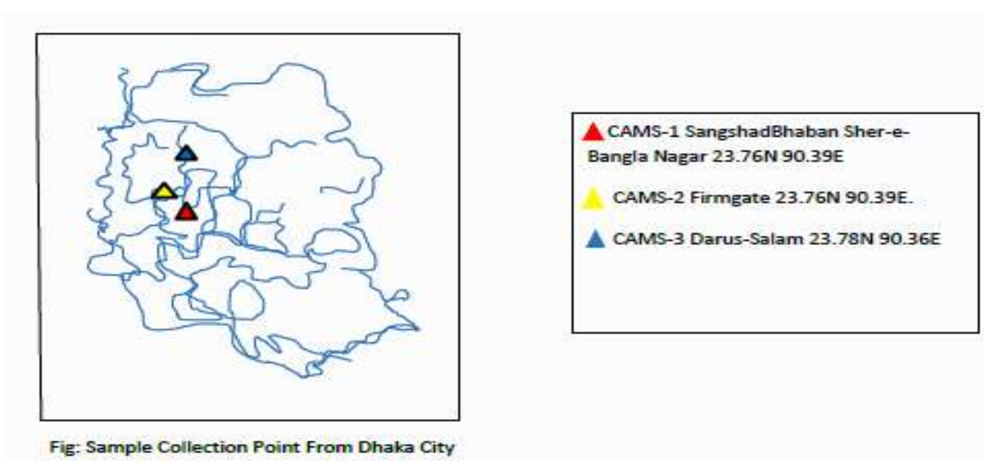

Fig. 5: Map location of air quality monitoring stations.

3.1 CAMS-01(Parliament Building Area, Dhaka)

The CAMS is working at the National Assembly premises at Manik Mia Avenue. Built in 2002, this is the principal consistent checking station in the nation. No nearby source near the station is present; in any case, significant city streets with moderate traffic thickness cross the territory with a separation of 500m. The inspecting bays are situated on the level top of the CAMS cover around 3 m over the ground and the admission spout is found 1.8 m over the rooftop with great regular ventilation.

3.2 CAMS-02 (BARC, Farmgate, Dhaka)

The second CAMS in Dhaka is situated at the first floor of the guardhouse of Bangladesh Agricultural Exploration Council (BARC) at Farmgate. The area is portrayed with high traffic territory with a convergence of a few principal streets with overwhelming traffic. The testing deltas are put on the level top of the CAMS cover. The rooftop stature is about 7 m over the ground and the admission spout of the sampler is found 1.8 m over the rooftop. The site is around 5 meter away from the significant traffic conduit with great common ventilation.

3.3 CAMS-03 (Darus Salam, Mirpur, Dhaka)

The third CAMS is arranged at the Mass Communication Establishment at Darus Salam zone. This area is moreover described with substantial traffic; an enormous number of vehicles from the northern piece of the nation enter the city along these lines. Significant block oven groups are likewise close to the site. The CAMS site is arranged around 100 meter away from the fundamental street. The rooftop tallness is around 7 m over the ground and the admission spout of the sampler is found 1.8 m over the rooftop.

Fig. 6: Air monitoring station (Parliament building area, Dhaka)

Fig. 7: Air monitoring station (BRRC, Firmgate, Dhaka)

Fig. 8: Air monitoring station (Darus Salam, Mirpur, Dhaka)

Table 2: Air quality of Dhaka city from 2013 to 2020 (μ = mean, M = median, SD = standard deviation).

Year	***SO_2***			***NO_2***			***O_3***			***CO_2***			***$PM_{2.5}$***			***PM_{10}***		
	μ	***M***	***SD***	***μ***	***M***	***SD***	***μ***	***M***	***SD***	***μ***	***M***	***SD***	***μ***	***M***	***SD***	***μ***	***M***	***SD***
2020	*6.7*	*6.5*	*2.7*	*34*	*33*	*12.7*	*6.7*	*7.1*	*1.8*	*1.9*	*2.1*	*0.35*	*154*	*139*	*59*	*240*	*214*	*74*
2019	*6.35*	*5.8*	*3.6*	*32.1*	*25.8*	*17.4*	*10.2*	*8.4*	*5.7*	*1.9*	*1.92*	*0.8*	*98*	*99*	*42.4*	*153*	*144*	*47*
2018	*10.5*	*6.5*	*9.32*	*54.5*	*53.3*	*17.5*	*5.8*	*5.7*	*2.1*	*1.8*	*1.6*	*1.1*	*98*	*76*	*64*	*163*	*129*	*110*
2017	*11.79*	*4.7*	*16.7*	*71.7*	*61*	*30.8*	*7.2*	*5.8*	*4.5*	*1.57*	*1.6*	*0.5*	*75*	*53*	*46*	*119*	*92*	*65*
2016	*7.5*	*6.1*	*4.6*	*61.4*	*66*	*29.8*	*4.5*	*4.6*	*2.3*	*1.5*	*1.3*	*0.6*	*81*	*53*	*57.2*	*146*	*98*	*90*
2015	*6.01*	*5.2*	*4.4*	*73.8*	*58.2*	*44.2*	*4.9*	*5*	*1.9*	*2.51*	*2.22*	*1.24*	*77.1*	*60*	*46*	*126*	*104*	*67*
2014	*5.06*	*3.14*	*4.34*	*60.2*	*41.5*	*50.6*	*8.6*	*6.8*	*3.8*	*1.7*	*1.6*	*0.7*	*80*	*63*	*56*	*136*	*125*	*81*
2013	*5.2*	*4.3*	*3.5*	*45.7*	*23.8*	*53.1*	*10.9*	*8.8*	*6.7*	*1.7*	*0.9*	*0.86*	*87*	*50*	*74*	*127*	*96*	*84*

4. METHODS AND METHODOLOGY

4.1 Autoregressive Moving Average (ARMA)

Autoregressive moving average (ARMA) model is an extensively used prediction model for forecasting. It is powerful tool for determining the behavior of time-series data. In autoregressive moving average (ARMA) model, it forecasts the future value using the linear relationship between the previous observed values. The parameters of ARMA model are the coefficients of linear relationship. The coefficients are based on the time-series data. By this way, it can easily forecast the future value by using the previous observed values. This computational compares to the training venture of the model, and the estimation of the approaching worth relates to the forecast step [18].

Autoregressive moving average (ARMA) model is a combination of p_{th}- order of autoregressive process and q_{th} – order of moving average. However, the ARMA (p,q) model can be formulated as:

$$Y_t = C + \sum_{i=1}^{p} \phi_i Y_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-1}$$

In the above equation (1), Y_t is the forecasted output, C is the internal constant, $\phi_1, \ldots, \phi_p$ are the autoregressive parameters and $\theta_1, \ldots, \theta_q$ are the moving average parameters.

4.2 Time Delay Neural Network (TDNN)

The time-delay neural network of action empowers the system to find acoustic-phonetic highlights and the temporal connections between them autonomous of position in time and henceforth not obscured by fleeting movements in the input [20]. The time-delay neural network architecture is designed for undermining speech sequences in an efficient manner. It is usually used for time-shift arrangement. It generates output regarding on input sequences. There are two extensive features for time-delay neural network.

1. The TDNN needs to perceive aftereffects that may happen at non-fixed situations in the input window. Subsequently the system needs to discover that the aftereffect is a component free of movements in its position.
2. The TDNN needs to perceive includes in any event, when those highlights show up at various relative positions. This circumstance emerges in situations where various aftereffects happen in the information window with various relative separations. This happens much of the time in genomic arrangements when at least one components are embedded or erased in a given advertiser [19].

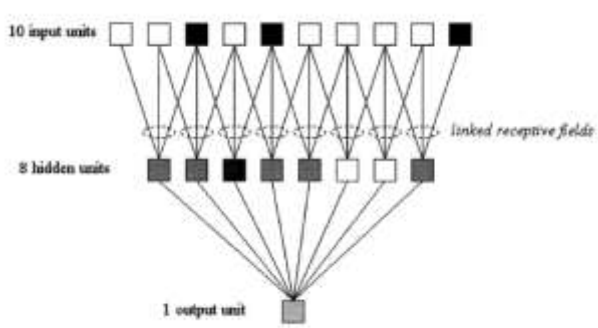

Fig. 9: The one layer time-delay neural network architecture network.

International Journal of Engineering and Information Systems (IJEAIS)
ISSN: 2643-640X
Vol. 4, Issue 8, August – 2020, Pages: 88-101

In TDNN there are 10 input units, some hidden layers are connected to the input units. The hidden layers have linked receptive fields. And it has an output unit. The receptive fields are responsive to the input window.

4.3 Long-Short Term Memory (LSTM)-Our Proposed Model

Deep learning models like Convolution Neural Network (CNN), Recurrent Neural Network (RNN) and Long Short Term Memory (LSTM) network have achieved great success recently and being widely used in video tracking, speech sequence modelling, image classification, emotion recognition and other fields due to their capability of self-learning and providing more accurate results. Recurrent Neural Network (RNN) has been applied to solve many problems namely speech recognition, language modelling, image classification [24]. However RNN has some drawbacks due to its vanishing and exploding gradient problems hence sometimes efficient outputs cannot be obtained in case of long-term temporal intervals [21]. Hence, Long Short Term Memory (LSTM) network which is a special type of RNN is introduced which has the ability to learn long temporal sequences and solve vanishing and exploding gradient problems [21].

In 1997 Hochreiter and Schmidhuber proposed an advanced version of RNN named as Long Short Term Memory (LSTM) network which has an improved structure than RNN and has the ability to learn long-temporal sequences [22]. The LSTM network consists of three gates: i) input gate ii) output gate iii) forget gate. These gates are added in the hidden layer of the network and thus it can resolve the problem of processing long temporal sequence of data [23].So LSTMs are specially designed to avoid the long temporal sequencing problem. Structure of LSTM network is illustrated in Fig. 8.

The first step of LSTM network is to determine which information should be kept and which information needs to be forgotten from the cell state. This is determined by a sigmoid layer named forget gate layer. The sigmoid layer looks at the current input $\mathbf{x_t}$ and last output $\mathbf{h_{t-1}}$ and outputs a number between **0** to **1** for each number in cell state $\mathbf{C_{t-1}}$. The output **1** refers that the data must be retained and the output **0** refers that the data must be eliminated. Thus we get a forget function:

$$f_t = \sigma(w_f.[h_{t-1},x_t] + b_f)$$

Here w_f is the weight matrices of the input layer and b_f is the corresponding bias vector.
The activation function σ (sigmoid) is characterized as $\sigma(x) = 1/(1+e^{-x})$.

The second step is to determine what new information is to be stored in the new cell state c_t. This cell state consist of two parts. In the first part a sigmoid layer named as "input gate layer" determines which values should be updated. In the second part a tanh layer appoints a new value of array c'_t that can be stored in the state. These two parts are then combined to update the state

$$i_t = \sigma(w_i.[h_{t-1},x_t] + b_i$$

$$c'_t = \tanh(w_c.[h_{t-1},x_t] + b_c)$$

Here $\mathbf{w_i}$ and $\mathbf{w_c}$ are the weight matrices of the input layer,b_i and b_c are the corresponding bias vectors. The activation function **tanh** is characterized as $tanh(x) = (e^x - e^{-x})/(e^x + e^{-x})$.

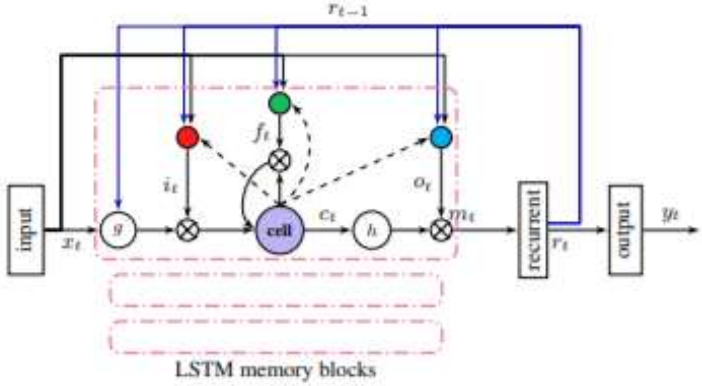

Fig. 10: LSTM model architecture [4]

The old cell state $\mathbf{c_{t-1}}$ is then upgraded to new memory $\mathbf{c_t}$. This is done by multiplying $\mathbf{i_t}$ with $\mathbf{c'_t}$ and then adding the term with the product of $\boldsymbol{f_t}$ and $\mathbf{c_{t-1}}$.

$$c_t = f_t * c_{t-1} + i_t * c'_t$$

Finally the output is determined based on the cell state. It is executed following a few steps. Firstly,a sigmoid layer is run to determine which part of the cell state should be returned as output. Then the cell state is assessed by the **tanh** function. It is then multiplied by the output of the sigmoid gate to obtain the filtered output. This stage can be expressed by the following equations:

$$o_t = \sigma(w_o.[h_{t-1},x_t] + b_o)$$

$$h_t = o_t * \tanh(C_t)$$

The output predicted value can be expressed by the following equation:

$$y_t = \sigma(w_y.h_t + b_y)$$

Here $\mathbf{w_o}$ and $\mathbf{w_y}$ are the weight matrices of the input layer, $\mathbf{b_o}$ and $\mathbf{b_y}$ are the corresponding bias vectors.

Apart from the traditional LSTM network some other variant LSTM networks are also being introduced by researchers for better optimization and to reduce the training time. One of the popular LSTM variant is "peephole connections" which was introduced by Gers & Schmidhuber in the year 2000.Another variation of LSTM network is to use coupled forget and input gates. Instead of separately deciding what to forget and what new information should be retained, this variant makes the decision together [24].Gated Recurrent Unit (GRU) is another LSTM variant. Unlike traditional LSTM structure that has three gates, GRU consists of two gates namely update gate and reset gate. This results in easy training, simple structure and fast weight gradient descent. Although it sometimes results in unstable training process [23].Some variants work better than others on certain tasks. However be it traditional or different variant LSTM networks, all serve the same purpose that is to work efficiently in case of long-term dependencies.

5. EXPERIMENTAL RESULTS AND DISCUSSIONS

Our overview on outdoor air pollution was to visualize the change in air quality from 2013 to 2020 in Dhaka city. This mega city is already renowned for being most air polluted in the world. The graph produced from the various air pollutants data is the prove that air quality of Dhaka city is degrading day by day, most probably for the high concentration of PM_{10} and $PM_{2.5}$. The increasing scenario of Industrial activities and fossil fuel burning, considering non-renewable resources as primary energy source is the consequences for this upcoming disaster.

The average amounts of $PM_{2.5}$ concentration from last 12 months data is 137.431 µg/m3 which is 9 times more than the national primary standard level. In 2013 this amount was 86.88917 µg/m3, the increase in7 years in about 1.6 times more. The PM_{10} scenario is worse, the average amounts of PM10 Conc. From last 12 months data is 194.0083 µg/m3, it is about 4 times more than the annual national primary standard level. In 2013 the average concentration was 127.9667 µg/m3. The other pollutants is still in tolerable limit according to the national primary standard level.

The daily (24-hr) average of unhealthy AQI for $PM_{2.5}$ is 150 µg/m3, if we see the monthly average almost every years Dec-March months data exceeds this limit due to seasonal variation and dry air. On the other hand similarly for PM_{10} it is 120-240 µg/m3, the pattern for PM_{10} is random. The most polluted time of the year is around Jan-Feb, the highest concentration was too in Jan 2020, almost 237 µg/m3. The highest concentration for PM_{10} was found during Feb 2020 which was around 420 µg/m3, this exceeds the hazardous level amount.

The data we collected can predict the future amounts and patterns for air pollutants in mega city like Dhaka. Also it can show the maximum protective measurement time and the minimum requirements throughout the year. The concern must be to minimize the dangerous elements first and keep the others in tolerable limit.

The amount of NO_2 was relatively lower in 2020 compare to 2013, the average amount in 2013 was about 45 ppb and the average in 2020 is 30 ppb. The data also shows that it was highest grossing in Dec'13. The average amount of NO_2 in approximately 7 years is 55 µg/m3 which is within the national primary level. Similarly CO_2, O_3, SO_2 amount also is within the limit so our main concern is to diminish the sources of particulate matter.

International Journal of Engineering and Information Systems (IJEAIS)
ISSN: 2643-640X
Vol. 4, Issue 8, August – 2020, Pages: 88-101

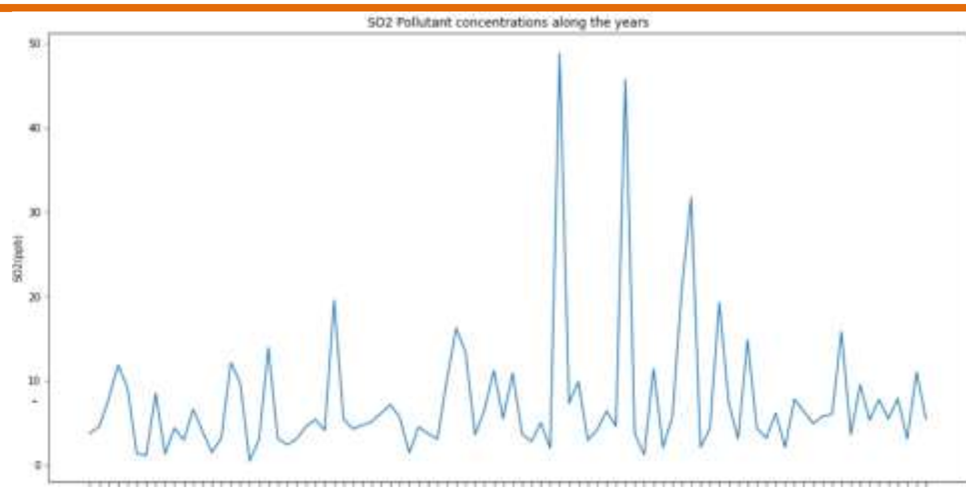

Fig. 11: SO_2 pollutant concentration (2013-2020)

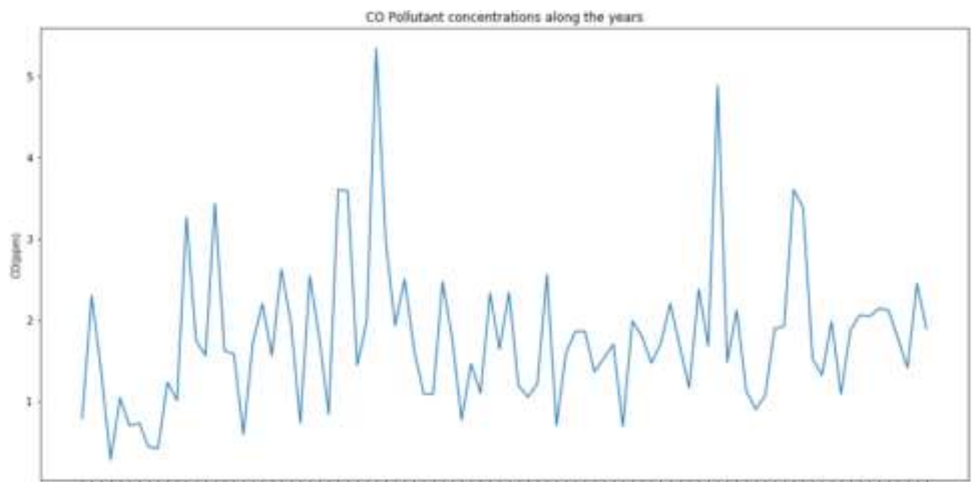

Fig. 12: CO_2 pollutant concentration (2013-2020)

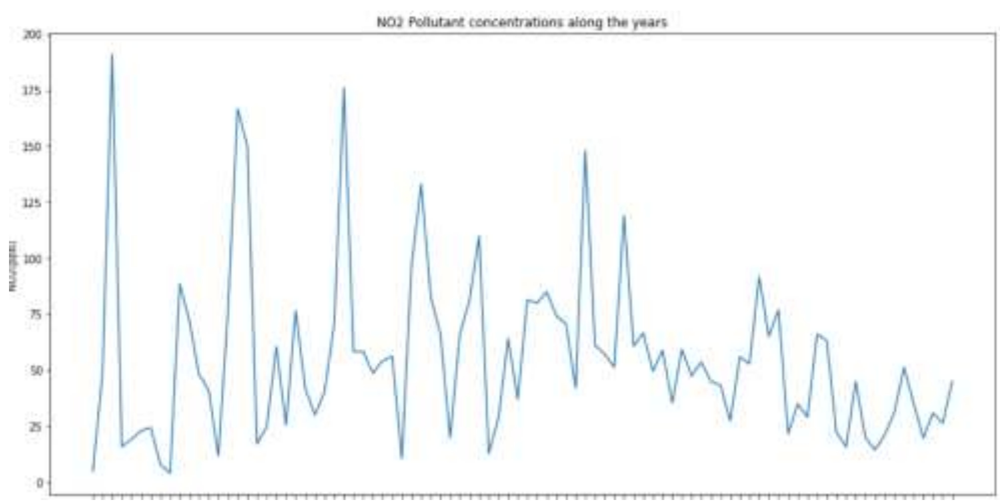

Fig. 13: NO_2 pollutant concentration (2013-2020)

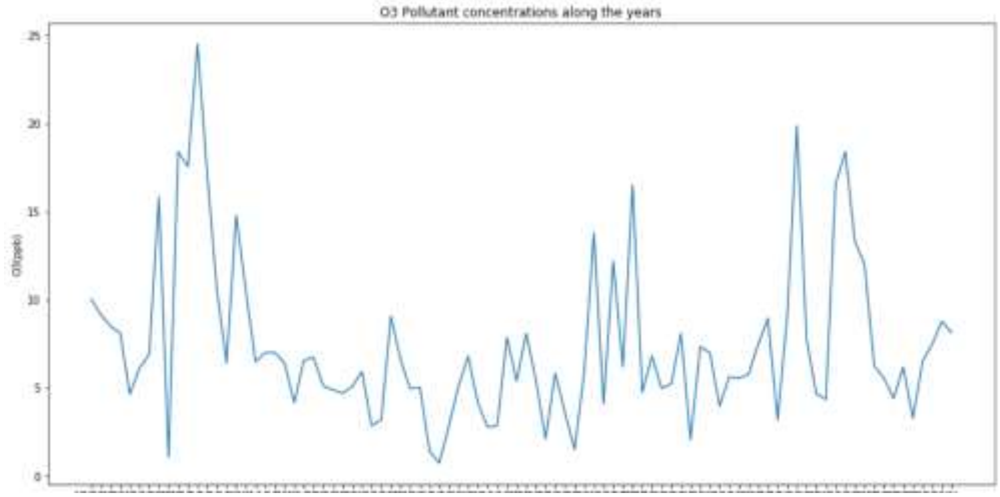

Fig. 14: O_3 pollutant concentration (2013-2020)

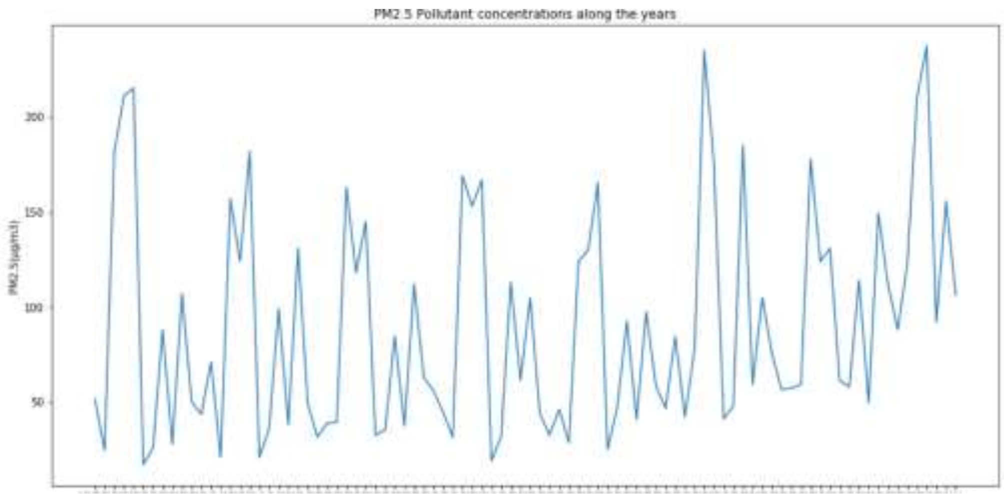

Fig. 15: $PM_{2.5}$ pollutant concentration (2013-2020)

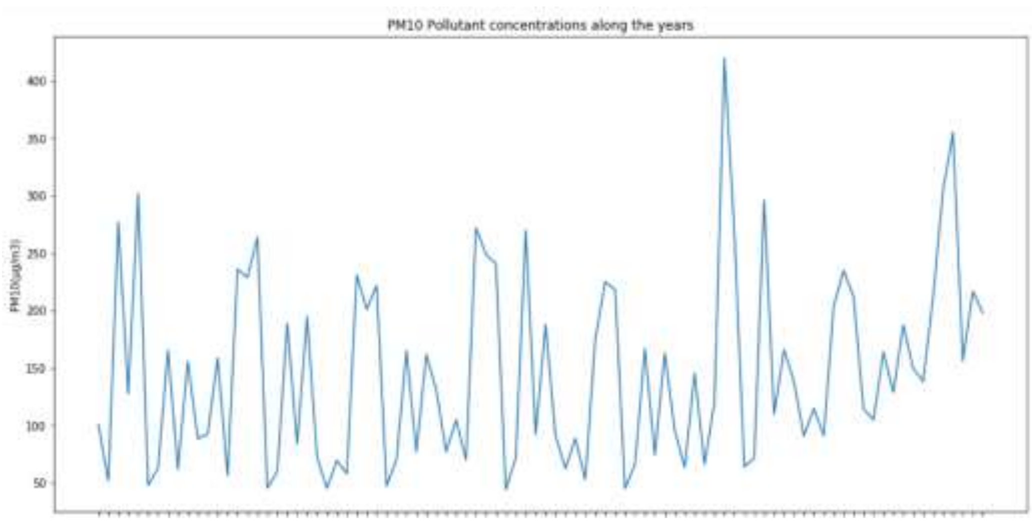

Fig. 16: PM_{10} pollutant concentration (2013-2020)

In this experiment, there are three extensively used models have been used. In our proposed model, long short term memory (LSTM), we split our dataset into test and train phase. We conducted three evaluation matrices such as root mean square error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE).

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - k_i)^2}$$

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|x_i - y_i|$$

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\frac{|x_i - y_i|}{y_i}$$

Here, y_i is the observed air pollutant concentration and x_i is the predicted air pollutant concentration, n is the number of samples.

In addition to this, we compared our model with other models such as ARMA and TDNN. These models were trained and tested using the same training and test sets applied for the LSTM model; however, the input data differed slightly in each model. The TDNN and ARMA models use the same inputs as our LSTM model, but the network architecture differs. Table 3 shows the different model performances. We can see form the table 3 that the RMSE value ranges from 9.23 to 28.26, the MAPE ranges from 10.27% to 20.57%. Our proposed LSTM model has performed outstandingly in terms of all evaluation matrices.

Table 3. Comparison of performances of different models.

Method	*RMSE*	*MAE*	*MAPE (%)*
LSTM (Proposed)	***9.23***	***4.41***	***10.27***
TDNN	*14.19*	*9.08*	*23.81*
ARMA	*28.26*	*14.06*	*29.57*

6. CONCLUSION

This paper represents an LSTM model to estimate the air pollutant concentration in Dhaka city, Bangladesh. Based on self-collected air pollutant dataset of Dhaka city from 2013 to June 2020. We also analyzed the impact of COVID-19 lockdown in Dhaka, Bangladesh. Our proposed LSTM model is ideal for time-series based analysis. It does not depend on long term data to predict more accurate outcome. But it can also do well in long term large dependency data. In this experiment, we evaluated our models in RMSE, MAPE and MAE to determine the best model for this purpose. However, daily SO_2, NO_2, CO_2, O_3, $PM_{2.5}$ and PM_{10} data have been collected from 3 different air quality monitoring stations in Dhaka, Bangladesh. In our study it confirms that LSTM based predicted has performed outstandingly. However, the RMSE value of LSTM is 9.23. Whereas MAE and MAPE values are 4.41 and 10.27% respectively. The other models ARMA and TDNN have very higher values than LSTM, where TDNN performed better than ARMA. So in this study, we have recommended LSTM model for further utilization in air pollution estimation.

REFERENCES

1. I. Manisalidis, E. Stavropoulou, A. Stavropoulos, and E. Bezirtzoglou, "Environmental and Health Impacts of Air Pollution: A Review," *Front. Public Heal.*, vol. 8, 2020.
2. D. Qin, J. Yu, G. Zou, R. Yong, Q. Zhao, and B. Zhang, "A Novel Combined Prediction Scheme Based on CNN and LSTM for Urban PM2.5 Concentration," *IEEE Access*, vol. 7, pp. 20050–20059, 2019
3. H. Akimoto and H. Akimoto, "Global Air Quality and Pollution Hajime Akimoto," vol. 1716, no. 2003, pp. 1716–1720, 2007.
4. J. O. Anderson, J. G. Thundiyil, and A. Stolbach, "Clearing the Air: A Review of the Effects of Particulate Matter Air Pollution on Human Health," *J. Med. Toxicol.*, vol. 8, no. 2, pp. 166–175, 2012.
5. B. Brunekreef and S. T. Holgate, "Air pollution and health," *Lancet*, vol. 360, no. 9341, pp. 1233–1242, 2002.
6. M. Kampa and E. Castanas, "Human health effects of air pollution," *Environ. Pollut.*, vol. 151, no. 2, pp. 362–367, 2008.

7. J. G. Lu, "Air pollution: A systematic review of its psychological, economic, and social effects," *Curr. Opin. Psychol.*, vol. 32, pp. 52–65, 2020.
8. M. Krishan, S. Jha, J. Das, A. Singh, M. K. Goyal, and C. Sekar, "Air quality modelling using long short-term memory (LSTM) over NCT-Delhi, India," *Air Qual. Atmos. Heal.*, vol. 12, no. 8, pp. 899–908, 2019.
9. C. M. Wong *et al.*, "Public Health and Air Pollution in Asia (PAPA): A multicity study of short-term effects of air pollution on mortality," *Environ. Health Perspect.*, vol. 116, no. 9, pp. 1195–1202, 2008.
10. T. Xayasouk, H. M. Lee, and G. Lee, "Air pollution prediction using long short-term memory (LSTM) and deep autoencoder (DAE) models," *Sustain.*, vol. 12, no. 6, 2020.
11. J. Kleine Deters, R. Zalakeviciute, M. Gonzalez, and Y. Rybarczyk, "Modeling PM2.5 Urban Pollution Using Machine Learning and Selected Meteorological Parameters," *J. Electr. Comput. Eng.*, vol. 2017, 2017.
12. T. Kitada and A. K. Azad, "Study on the air pollution control system for dhaka, bangladesh," *Environ. Technol. (United Kingdom)*, vol. 19, no. 5, pp. 443–459, 1998.
13. B. A. Begum and P. K. Hopke, "Ambient air quality in dhaka bangladesh over two decades: Impacts of policy on air quality," *Aerosol Air Qual. Res.*, vol. 18, no. 7, pp. 1910–1920, 2018.
14. M. F. Demirbas, K. Bozbas, and M. Balat, "Carbon dioxide emission trends and environmental problems in Turkey," *Energy Explor. Exploit.*, vol. 22, no. 5, pp. 355–366, 2004.
15. S. J. Davis, K. Caldeira, and H. D. Matthews, "Future CO2 emissions and climate change from existing energy infrastructure," *Science (80-.).*, vol. 329, no. 5997, pp. 1330–1333, 2010.
16. T. Boningari and P. G. Smirniotis, "Impact of nitrogen oxides on the environment and human health: Mn-based materials for the NOx abatement," *Curr. Opin. Chem. Eng.*, vol. 13, no. x, pp. 133–141, 2016.
17. L. Watson *et al.*, "Impact of emissions and +2 °C climate change upon future ozone and nitrogen dioxide over Europe," *Atmos. Environ.*, vol. 142, pp. 271–285, 2016.
18. P. Gomes and R. Castro, "Wind Speed and Wind Power Forecasting using Statistical Models: AutoRegressive Moving Average (ARMA) and Artificial Neural Networks (ANN)," *Int. J. Sustain. Energy Dev.*, vol. 1, no. 2, pp. 41–50, 2012.
19. M. G. Reese, "Application of a time-delay neural network to promoter annotation in the Drosophila melanogaster genome," vol. 26, pp. 51–56, 2001.
20. K. Shikano and K. J. Lang, "Phoneme Recognition Using Time-Delay Neural Networks I "," vol. 31, no. 3, pp. 328–339, 1989.
21. B. Nakisa, M. N. Rastgoo, and A. Rakotonirainy, "Long Short Term Memory Hyperparameter Optimization for a Neural Network Based Emotion Recognition Framework," *IEEE Access*, vol. 6, pp. 49325–49338, 2018.
22. D. H. Nguyen-le, Q. B. Tao, V. Nguyen, and M. Abdel-wahab, "A data-driven approach based on long short-term memory and hidden Markov model for crack propagation prediction," *Eng. Fract. Mech.*, p. 107085, 2020.
23. S. Pei, H. Qin, Z. Zhang, L. Yao, Y. Wang, and C. Wang, "Wind speed prediction method based on Empirical Wavelet Transform and New Cell Update Long Short-Term Memory network," *Energy Convers. Manag.*, vol. 196, no. June, pp. 779–792, 2019.
24. F. Hu, Y. Zhu, J. Liu, and L. Li, "An efficient Long Short-Term Memory model based on Laplacian Eigenmap in artificial neural networks," *Appl. Soft Comput. J.*, vol. 91, p. 106218, 2020.