Journal of Korea Multimedia Society Vol. 25, No. 2, February 2022(pp. 440-449)
https://doi.org/10.9717/kmms.2022.25.2.440

Exploiting Neural Network for Temporal Multi-variate Air Quality and Pollutant Prediction

Muneeb A. Khan[†], Hyun-chul Kim[††], Heemin Park[†††]

ABSTRACT

In recent years, the air pollution and Air Quality Index (AQI) has been a pivotal point for researchers due to its effect on human health. Various research has been done in predicting the AQI but most of these studies, either lack dense temporal data or cover one or two air pollutant elements. In this paper, a hybrid Convolutional Neural approach integrated with recurrent neural network architecture (CNN-LSTM), is presented to find air pollution inference using a multivariate air pollutant elements dataset. The aim of this research is to design a robust and real-time air pollutant forecasting system by exploiting a neural network. The proposed approach is implemented on a 24-month dataset from Seoul, Republic of Korea. The predicted results are cross-validated with the real dataset and compared with the state-of-the-art techniques to evaluate its robustness and performance. The proposed model outperforms SVM, SVM-Polynomial, ANN, and RF models with 60.17%, 68.99%, 14.6%, and 6.29%, respectively. The model performs SVM and SVM-Polynomial in predicting O3 by 78.04% and 83.79%, respectively. Overall performance of the model is measured in terms of Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) and the Root Mean Square Error (RMSE).

Key words: Air Quality Index (AQI), Air Pollutant, Temporal clustering, Forecasting

1. INTRODUCTION

Air is one of the most basic elements for human survival and good air quality is necessary for human health. Unfortunately, as a consequence of the industrialization and urbanisation era, the air quality deteriorates constantly. Nowadays, air pollution has become a global problem, which has aroused widespread concern from scholars, governments and the public. According to the World Health Organisation (WHO), each year more than 1.3 million people worldwide die prematurely due to low air quality [1]. Furthermore, long-term exposure to such areas is associated with the occurrence of many diseases such as respiratory disease, cardiovascular disease and even cancer, contributing to as many as 4 - 9 million human deaths per year globally [2,3,4].

In order to tackle this issue, countries monitor the air quality and design plans to address this issue. The Environmental Protection Agency (EPA) commonly monitors carbon monoxide (CO), fine particulate matter ($PM_{2.5}$), respirable particulate matter (PM_{10}), sulphur dioxide (SO_2), nitrogen dioxide (NO_2) and ozone (O_3) to determine air quality, commonly known as Air Quality Index (AQI). AQI indicates whether the air is clear, polluted or even hazardous in that specific region. Table 1 indicates the EPA standard of under-observation elements and their range.

※ Corresponding Author : Heemin Park, Address: (31066) 31 Sangmyungdae-gil, Dongnam-gu, Cheoan, South Korea, TEL : +82-41-550-5367, FAX : +82-41-550-5386, E-mail : heemin@smu.ac.kr
Receipt date : Feb. 9, 2022, Approval date : Feb. 17, 2022

[†] Dept. of Software, Sangmyung University
(E-mail : muneebkhan046@gmail.com)
[††] Dept. of Software, Sangmyung University
(E-mail : hkim@smu.ac.kr)
[†††] Dept. of Software, Sangmyung University
※ This research was funded by a 2020 research Grant from Sangmyung University, South Korea.

Table 1. The Environmental Protection Agency (EPA) Air Pollutant Standard.

	CO (ppm)	$PM_{2.5}$ (μg/m^3)	PM_{10} (μg/m^3)	SO_2 (ppm)	NO_2 (ppm)	O_3 (ppm)
Good	2	15	30	0.02	0.03	0.03
Normal	9	35	80	0.05	0.06	0.09
Bad	15	75	150	0.15	0.2	0.15
Hazardous	50	500	600	1	2	0.5

Recently, sensors and actuators are used to monitor the air quality because of its manifold advantages such as cost-effectiveness, ease to deploy and monitoring, remote access and accumulation of the data. This collection of data can be very helpful for researchers and scientists to monitor, analyse and predict using data-driven approaches.

Data-driven approaches use geo-temporal, non-linear data and with the help of Artificial Intelligence (AI), statistical and mathematical analysis provide an in-depth of problems and look for the solution to mitigate the problems [5, 6]. The exploitation of various techniques such as Long short-term memory (LSTM), Random Forest (RF), Support Vector Machine (SVM), K-mean, Decision Tree, neuro Fuzzy techniques are used to efficiently predict the air pollutant and AQI [9,12]. Author of [7,8] used Convolutional Neural Network (CNN) to calculate and predict the respirable particulate matter (PM_{10}). A hybrid CNN based approach is proposed to minimise the prediction error of O_3 [13].

In recent years, a lot of research has been done in forecasting air pollutants and AQI by either exploiting CNN or AI techniques. However, they often forecast just one or two pollutant elements, lack a dense dataset, the correlation between air pollutants. There is a necessity for an unbiased, robust model which monitor and forecast multivariate (EPA monitored) air pollutant elements. In this paper, we have presented a hybrid CNN-LSTM based approach for AQI and multiple air pollutants prediction. We opted for CNN due to its efficient feature extraction and high processing capability of grid data. The integration of CNN with LSTM provides a unique ability to extract features using CNN while LSTM; a recurrent neural network, associates multi-variate and their interdependencies.

The main contributions of this paper are as follows:

1. A CNN-LSTM amalgamated model is presented for multivariate correlation and forecasting of air pollutant elements. Furthermore, the predictions are validated with more than 10 days of real-time acquired air pollutant data.

2. Most of the previous research is mostly focused on uni or bi-variate i.e PM_{10} & $PM_{2.5}$. In this paper, we use multivariate pollutant elements such as CO, $PM_{2.5}$, PM_{10}, SO_2, NO_2 and O_3.

3. An in-depth statistical, geo-temporal and temporal analysis is presented.

4. To the best of our knowledge, this is the first time a study of these multivariate variables has been implemented in the city of Seoul, Republic of Korea.

5. Design and development of the model which governs, retrieves and predicts the critical information of multivariate pollutant elements. Additionally, the model is rigorously evaluated and results are compared with the state-of-the-art models.

6. A detailed study on the data of the 24 months period is performed and findings are discussed. Furthermore, cross-validation of the predicted results is done with the real-time calculated results.

The rest of the paper is as follows. Section 2 discussed the state of the art and recently published research paper in the said domain. In section 3, we converse about the methodology, dataset, pre-processing and transformation. Section 4 confers about results and a wide-range discussion is also presented on the acquired results. A conclusion is presented in section 5.

2. RELATED WORK

In this section, we discussed the research, state of the art techniques that have been published in this domain.

Authors of [5], presented a graph neural network-based approach, called Attentional Deep Air quality Inference Network (ADAIN), for urban AQI prediction. ADAIN employs an attention base layer to extract and learn the weight and relation between pollutant elements to improve the overall prediction performance. The proposed model presents the RMSE of 42, 33 and 25 for $PM_{2.5}$, PM_{10}, and NO_2 respectively. Furthermore, an accuracy of 62%, 71% and 79% has been achieved to accurately predict the $PM_{2.5}$, PM_{10}, and NO_2.

In [9] multi-regression method has been implemented for AQI analysis on the dataset of USA and India cities. Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RSME) has been selected to evaluate the accuracy of proposed techniques. The proposed, multi-linear regression, model achieves an MAE, MAPE and RMSE of 5.13%, 3.01% and 6.2% respectively.

The author of [10], exploits Multinomial Logistics Regression (MLR), Decision Tree and K mean techniques for efficient AQI prediction. The author classifies the result into Good, Normal and Hazardous . MLR outperformed the Decision Tree technique with an error rate of 0.44 in comparison to 0.66 respectively.

In [14], authors proposed a CNN base $PM_{2.5}$ monitoring system using natural image. Hybrid CNN and feature based CNN-RF model is presented to classify the images with respect to $PM_{2.5}$. The proposed models have achieved an accuracy of 68.74% in comparison to RF with a classification accuracy of 63.62%. Authors of [15] proposed a 1-dimensional CNN with Exponential Adaptive Gradients (EAG) to monitor 4 air pollutant elements (PM_{10}, CO, O_3 and NO_2). The proposed model maintains the history of gradient to adjust the learning rate and check points are used to update the weights. The proposed model achieved an overall error rate of MAPE, MAE and R^2 with the 2.35, 4.214 and 0.996 respectively.

In [16], a hybrid framework is presented which can cope with hierarchical multi-factor temporal dataset to predict $PM_{2.5}$. The author integrates 1-dimensional CNN with bi-directional LSTM to extract features from multivariate time series and interrelation between them. The proposed model is unable to adjust with respect to anomalies or sudden change in $PM_{2.5}$ level.

In [17], the authors presented a hybrid prediction model to efficiently predict AQI for the Thailand region. Genetic programming based MLR model has been suggested for bi-class (healthy and unhealthy) AQI prediction. The proposed model acquired an accuracy of 97%, 80.7% for healthy and unhealthy AQI. Moreover, this study is performed on AQI instead of air pollutants like PM_{10}, SO_2, NO_2 and O_3.

The authors of [18] proposed a hybrid, Autoregressive Integrated Moving Average Time Series, model to examine and forecast the $PM_{2.5}$ for short term prediction in Beijing, China. The proposed model acquired an RMSE of 26.6.

In [19], Auto Regression models are evaluated to predict the AQI for a week. An hourly dataset of 11 days has been used for training and testing the performance of multi regression models. The dataset is divided into 73 to ratio 27 for training and testing respectively. This study is carried out on 4 pollutants ($PM_{2.5}$, PM_{10}, SO_2 and NO_2). The proposed model has acquired an overall RMSE of 27.

In [20], authors presented a heterogeneous neural network; Multi-adversarial spatiotemporal recurrent Graph Neural Networks (MasterGNN), to predict AQI and weather efficiently. The proposed model is trained on Beijing, China and Shanghai, China air quality dataset. Furthermore, Master GNN exploits the spatiotemporal data and its dependency on weather. Authors of [21], proposed a stochastic optimization model base algorithm called Particle Swarm Optimization-Back Propagation (PSO) to predict the AQI. Author integrates a hybrid Back Propagation neural network with PSO (IPSO-BP), to get improved and better results. The study is carried out on 4 pollutant elements $PM_{2.5}$, PM_{10}, SO_2 and NO_2. IPSO-BP generates an overall error of 0.0326 and 0.072 for training and testing samples respectively.

3. METHODOLOGY

3.1 Dataset

The dataset used in this study is acquired from Seoul, Republic of Korea online data plaza. The acquired dataset is collected from 25 different locations in Seoul and from 01 January 2017 to 31 December 2019. The original dataset has many columns such as Station Number, Station Address, Longitude, Latitude, CO, $PM_{2.5}$, PM_{10}, SO_2, NO_2 and O_3. The dataset contains more than 600,000 instances of hourly averaged air pollutant elements collected daily. In Seoul, daily 3600 air quality instances are calculated from all over the city. A generic matrix, AQI_{Seoul}, is designed which depict the mathematical model of the acquired dataset as shown below:

In $AQI_{Seoul,}$ x represents the overall multiple parameters such as station code, Address, Latitude, Longitude, CO, $PM_{2.5}$, PM_{10}, SO_2, NO_2 and O_3, which are calculated after regular interval of time T where $\{T \in y : 1 \le y \le 24\}$. Furthermore, M represents the total number of stations where $\{M \in i : i = 1, 2, ..., n-1, n\}$. In pursuance of better accuracy and AQI prediction air pollutant elements (CO, $PM_{2.5}$, PM_{10}, SO_2, NO_2 and O_3) are studied without considering any dependencies. The acquired dataset has more than 600,000 instances as shown in Table 2.

3.2 Data Pre-Processing

Noise, irregularities and outliers in the dataset can directly affect the output and learning phase of the model. Henceforth, before forwarding the data to the model for the training phase, pre-processing techniques such as normalisation, outlier identifying, cleansing and removing of duplication is performed. Furthermore, to create a meaningful relation and uniformity, the acquired dataset is processed.

A total of 4992 anomalies or irregular instances such as values less than or equal to 0 are identified (as shown in Fig. 1 a & b) from the dataset. In order to create a consistency among the dataset, all those instances containing irregular values are removed from the dataset.

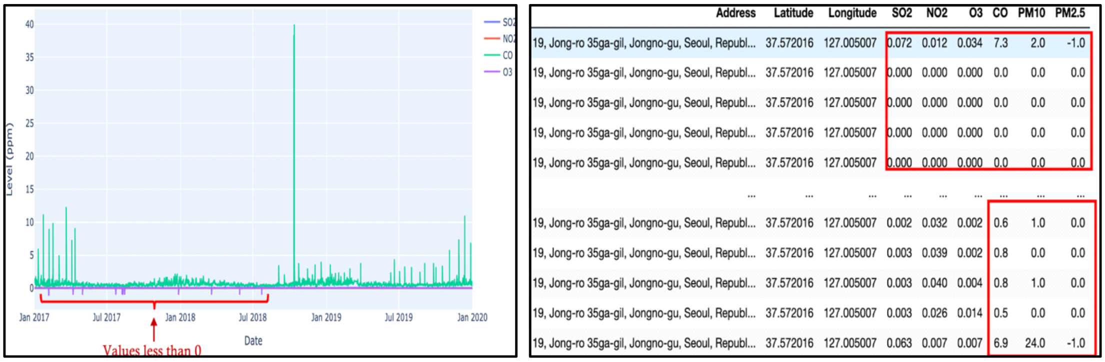
3.3 Proposed Model

In this paper we use a hybrid model by integrating CNN with LSTM. The purpose of opting for LSTM is due to its salient feature of the memory cell (as shown in Fig. 2a) which can save historical information. Moreover, CNN-LSTM, essentially incorporates the features of CNN with LSTM, which has generated some ground-breaking results, particularly in natural language processing and pattern identification and classification. The

$$AQI_{Seoul} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{M-1} \\ x_M \end{bmatrix} = \begin{bmatrix} X_1^{T1} & X_1^{T2} & X_1^{T3} & & X_1^{T22} & X_1^{T23} & X_1^{T24} \\ X_2^{T1} & X_2^{T2} & X_2^{T3} & & X_2^{T22} & X_2^{T23} & X_2^{T24} \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ X_{n-1}^{T1} & X_{n-1}^{T2} & X_{n-1}^{T3} & & X_{n-1}^{T22} & X_{n-1}^{T23} & X_{n-1}^{T24} \\ X_n^{T1} & X_n^{T2} & X_n^{T3} & & X_n^{T22} & X_n^{T23} & X_n^{T24} \end{bmatrix} \quad (1)$$

Table 2. Original Dataset.

Date	Station code	Address	Latitude	Longitude	SO_2	NO_2	O_3	CO	PM_{10}	$PM_{2.5}$
2017/01/01	101	19, Jong-ro 35ga-gil, Jongno-gu, Seoul, Korea	37.572016	127.005007	0.004	0.059	0.002	1.2	73	57
2017/01/02	101	19, Jong-ro 35ga-gil, Jongno-gu, Seoul, Korea	37.572016	127.005007	0.004	0.058	0.002	1.2	71	59
2017/01/03	101	19, Jong-ro 35ga-gil, Jongno-gu, Seoul, Korea	37.572016	127.005007	0.004	0.056	0.002	1.2	70	59
...	...	...	...	...	...	...	...	...	...	...
2019/12/29	125	59, Gucheonmyeon-ro 42-gil, Gangdong-gu, Seoul, Korea	37.544962	127.136792	0.003	0.023	0.015	0.4	24	17
2019/12/30	125	59, Gucheonmyeon-ro 42-gil, Gangdong-gu, Seoul, Korea	37.544962	127.136792	0.003	0.04	0.004	0.5	25	18
2019/12/31	125	59, Gucheonmyeon-ro 42-gil, Gangdong-gu, Seoul, Korea	37.544962	127.136792	0.003	0.037	0.005	0.5	27	18



Address	Latitude	Longitude	SO2	NO2	O3	CO	PM10	PM2.5
19, Jong-ro 35ga-gil, Jongno-gu, Seoul, Republ...	37.572016	127.005007	0.072	0.012	0.034	7.3	2.0	-1.0
19, Jong-ro 35ga-gil, Jongno-gu, Seoul, Republ...	37.572016	127.005007	0.000	0.000	0.000	0.0	0.0	0.0
19, Jong-ro 35ga-gil, Jongno-gu, Seoul, Republ...	37.572016	127.005007	0.000	0.000	0.000	0.0	0.0	0.0
19, Jong-ro 35ga-gil, Jongno-gu, Seoul, Republ...	37.572016	127.005007	0.000	0.000	0.000	0.0	0.0	0.0
19, Jong-ro 35ga-gil, Jongno-gu, Seoul, Republ...	37.572016	127.005007	0.000	0.000	0.000	0.0	0.0	0.0
...	...	...	...	...	...	...	...	...
19, Jong-ro 35ga-gil, Jongno-gu, Seoul, Republ...	37.572016	127.005007	0.002	0.032	0.002	0.6	1.0	0.0
19, Jong-ro 35ga-gil, Jongno-gu, Seoul, Republ...	37.572016	127.005007	0.003	0.039	0.002	0.8	0.0	0.0
19, Jong-ro 35ga-gil, Jongno-gu, Seoul, Republ...	37.572016	127.005007	0.003	0.040	0.004	0.8	1.0	0.0
19, Jong-ro 35ga-gil, Jongno-gu, Seoul, Republ...	37.572016	127.005007	0.003	0.026	0.014	0.5	0.0	0.0
19, Jong-ro 35ga-gil, Jongno-gu, Seoul, Republ...	37.572016	127.005007	0.063	0.007	0.007	6.9	24.0	-1.0

Fig. 1. Data Anomalies in the Dataset.

LSTM's specific formula derivation is as follows:

$$f_t = \sigma \left(W_f \cdot [h_{t-1}, x_t] + bf \right) \tag{2}$$

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

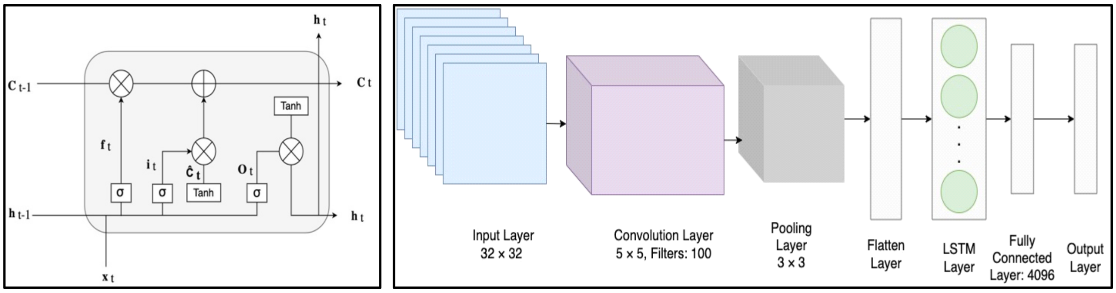
$$C_t = tanh \left(W_C \cdot [ht_{-1}, x_t] + b_C \right) \tag{4}$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot C_t \tag{5}$$

$$o_t = \sigma \left(W_o \cdot [h_{t-1}, x_t] + b_o \right) \tag{6}$$

$$h_t = o_t \cdot tanh(Ct) \tag{7}$$

This paper aim to design a hybrid deep learning model with BP time to obviate the vanishing gradient problem and forecast a city-wide AQI. The proposed framework uses 5×5 CNN convolutional layer to extract the various features such as multivariate air pollutant elements, location and time series. The 5×5 convolutional layer is opted to improve the learning of model and find the interrelationship between the air pollutant elements and time series. Moreover, convolutional and pooling layers of CNN are used to extract relevant features (as shown in Fig. 2b). The extracted features converted into a 1-D array and were forwarded to the LSTM as an input for model training. The LSTM act as a temporal sequence analyzer to analyze the temporal order of the said input while keeping its



(a) (b)

Fig. 2. Detail Structure of Proposed Model. (a) Memory Cell of LSTM and (b) Structure of Proposed Model.

state hidden throughout the training. In the end, AQI of the nth hour is predicted for CO, $PM_{2.5}$, PM_{10}, SO_2, NO_2 and O3 respectively.

3.3.1 Model Training

Before training, the dataset is divided into 2 parts; Training and Testing, consequently. Out of 642,519 collected AQI samples, 640,00 samples have been used for training while 2,519 samples are used for the testing phase. Predicted AQI results are validated by comparing them with real AQI samples results. The opted test samples are from 19 December 2019 to 31 December 2019. Furthermore, RMSE, MAPE and MAE were calculated to error rate, performance and overall efficiency of the proposed model.

The opted test samples are from 19 December 2019 to 31 December 2019. The aim of validating predicted results with real dataset acquired values is to design a reliable, robust system and check the overall performance of the proposed model. Henceforth, 2519 test samples are selected for cross-validation and testing the predicted results.

3.3.2 Experimental Setup

We implemented the model in Keras on Tensorflow using Adam optimizer with a batch size of 1024 for model training. Experimentation is carried out on the Jupyter Notebook with a GeForce RTX 2080 Ti GPU. An overall view of the experimental environment is shown in Table 3.

Table 3. Experimental Setup Overview.

Experimental Environment	Details
GPU	GeForce RTX 2080 Ti
Operating System	Ubuntu
CUDA	11.2
Python	3.6
Optimizer	Adam

4. RESULTS AND DISCUSSIONS

This section is organized into two parts. In the first part, a temporal analysis of the dataset is discussed. A generalized discussion of the overall air pollutant elements over the period of 2017 to 2019 is presented. However, in the second part, the results of the proposed model, its prediction and a comparison analysis are conferred.

4.1 Temporal Analysis of Air Pollutant Elements

In the temporal dimension, the annual air pollutant elements; CO, $PM_{2.5}$, PM_{10}, SO_2, NO_2 and O_3, are analyzed using the available dataset. The time series analysis is performed on the Seoul AQI dataset from January 2017 to December 2019 (as shown in Fig. 3). On average the value SO_2, NO_2, CO, O_3, PM_{10}, and $PM_{2.5}$ of during this period was 0.00436, 0.0286, 0.518, 0.024, 43.981 and 25.5689 respectively.

In yearly and seasonal basis analysis, it has been observed that the values of SO_2 and NO_2 remain within range and below the Bad threshold while

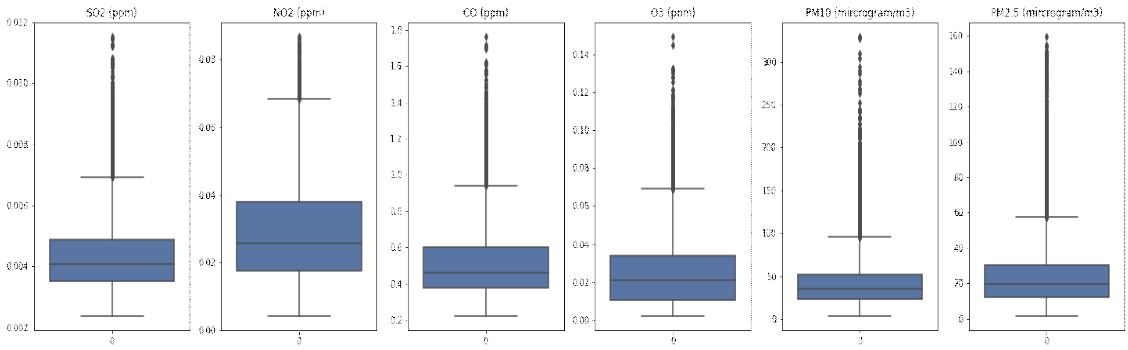


Fig. 3. Distribution of Air Pollutants in Seoul (2017-2019).

on the other hand the CO, O_3, PM_{10}, and $PM_{2.5}$ fluctuate and goes beyond the hazardous level. During winter the values of PM_{10}, $PM_{2.5}$ reach the value of 388, 160 respectively.

4.2 AQI Prediction Model

In this paper, a hybrid CNN-LSTM model is implemented to forecast the AQI of Seoul, the Republic of Korea with an aim to efficiently monitor the air quality. In order to check the robustness of the model, we use MAE, MAPE and RMSE as an evaluation parameters. Moreover, the proposed model predicted 12 days results (19 December to 31 December 2019) for each pollutant element. Acquired predicted results are cross validated with the real data of the said time frame and error rates for each air pollutant element are calculated respectively. The aim is to provide an unbiased, reliable and robust model. Predicted model results are illustrated in Table 4.

A comparison of the proposed model with the state-of-the-art model is performed to evaluate the performance of the presented model (as shown in Fig. 4). We select the SVM, SVM-Polynomial, ANN, KNN and RF to compare the results of our model.

Table 4. Results of AQI Prediction in terms of Error.

	MAE	RMSE	MAPE
SO_2	0.00140	3.05	0.009
NO_2	0.0123	15	0.37
CO	0.139	10.49	0.2245
O_3	0.0117	2.013	3.82
PM_{10}	9.11	17.463	0.227
$PM_{2.5}$	8.759	20.564	0.270

The proposed models perform better in comparison with other state-of-the-art models. Proposed model outperforms the SVM, SVM-Polynomial, ANN and RF model in forecasting SO_2 with 60.17 %, 68.99%, 14.6% and 6.29% respectively. Overall, the model outperforms SVM, SVM-Polynomial with 78.04% and 83.79% better results in forecasting O_3.

In this paper, we implemented the ANN model with Multilayer Perceptron (MLP) integrated with BP to evaluated and compare the model performance with the proposed CNN-LSTM model. The number of hidden layers were adjusted as flexible and adjusted until the best optimal results are acquired. The number of input neuron is 7 containing air pollutant elements (CO, $PM_{2.5}$, PM_{10}, SO_2, NO_2 and O_3) and times series. However, the proposed still able to outperform the ANN for CO, $PM_{2.5}$, PM_{10}, SO_2, NO_2 and O_3 with the 64%, 33.49%, 14.6%, 14.613%, 20.15% and 48.5% respectively.

5. CONCLUSIONS

Air is one of the fundamental elements for human survival. However, with rapid industrialization and urbanization, the quality of air suffers a

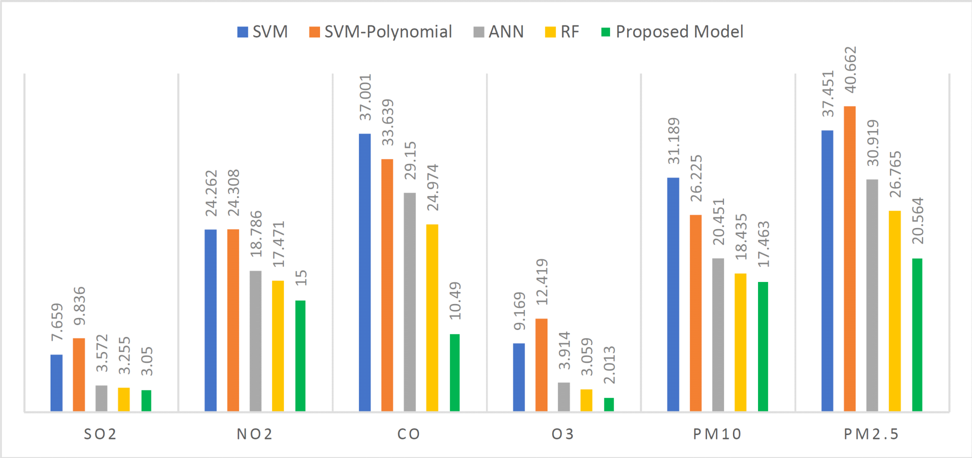


Fig. 4. Comparison Analysis of Proposed Model.

drastic downfall which causes cardiovascular and respiratory diseases. These air pollutants are not only threat human health to but also a major threat to ecosystem. Hence, air pollution has become a global problem, which has aroused widespread concern from research scholars, governments and the public. In recent years, a lot of research has been done but often they are done on univariate or bivariate air pollutant elements. In this paper, we presented a Deep Learning model to predict air quality by leveraging monitoring station data from 2017 to 2019. We opted CNN-LSTM base model due to the salient feature of efficient feature extraction and perseverance of historical information. Furthermore, a temporal analysis of air quality in Seoul, the Republic of Korea from 2017 to 2019 on various air pollutants elements such as CO, $PM_{2.5}$, PM_{10}, SO_2, NO_2 and O_3. RMSE, MAE and MAPE choose to evaluate model performance. The results demonstrate that the proposed model generate promising results in term of efficiency and forecasting AQI. In future, we explore an intercorrelation between AQI and weather index and evaluate various cities datasets.

REFERENCE

[1] World Health Organization, Air pollution, https://www.who.int/westernpacific/health-topics/air-pollution (Accessed February 7, 2022).

[2] X. Li, L. Jin, and H. Kan, "Air Pollution: a Global Problem Needs Local Fixes," *Nature*, Vol. 570(7762), pp. 437-439, 2019.

[3] B. Gardiner, *Choked: Life and Breath in the Age of Air Pollution*, University of Chicago Press: Chicago, IL, USA, 2019.

[4] G.H. Jang, H.H. Park, S.L. Lee, D.H. Kim, and M.K. Lim, "An Effective Extraction Algorithm of Pulmonary Regions Using Intensity-level Maps in Chest X-ray Images," *Journal of Korea Multimedia Society*, Vol. 13, No. 7, pp. 1062-1075, 2010.

[5] W. Cheng, Y. Shen, Y. Zhu, and L. Huang, "A Neural Attention Model for Urban Air Quality Inference: Learning the Weights of Monitoring Stations," *Thirty-second AAAI Conference on Artificial Intelligence*, pp. 2151-2158, April, 2018.

[6] Y. Rybarczyk and R. Zalakeviciute, "Machine Learning Approaches for Outdoor Air Quality Modelling: A Systematic Review," *Applied Sciences*, Vol. 8, Issue 12, p. 2570, 2018.

[7] J.M. Garcia, F. Teodoro, R. Cerdeira, L.M.R. Coelho, P. Kumar, and M.G. Carvalho, "Developing a Methodology to Predict PM10

Concentrations in Urban Areas using Generalized Linear Models," *Environmental Technology*, Vol. 37, Issue 18, pp. 2316-2325, 2016.

[8] S. Park, M. Kim, M. Kim, H.G. Namgung, K.T. Kim, K.H. Cho and S.B. Kwon, "Predicting PM10 Concentration in Seoul Metropolitan Subway Stations using Artificial Neural Network (ANN)," *Journal of Hazardous Materials*, Vol. 341, pp. 75-82, 2018.

[9] S.S. Ganesh, S.H. Modali, S.R. Palreddy, and P. Arulmozhivarman, "Forecasting Air Quality Index using Regression Models: A Case Study on Delhi and Houston," *International Conference on Trends in Electronics and Informatics (ICEI)*, pp. 248-254, 2017.

[10] K. Nandini and G. Fathima, "Urban Air Quality Analysis and Prediction Using Machine Learning," *1st International Conference on Advanced Technologies in Intelligent Control, Environment, Computing & Communication Engineering (ICATIECE)*, pp. 98-102, 2019.

[11] M. Alizamir, O. Kisi, R.M. Adnan, and A. Kuriqi, "Modelling Reference Evapotranspiration by Combining Neuro-fuzzy and Evolutionary Strategies," *Acta Geophysica*, Vol. 68, pp. 1113-1126, 2020.

[12] Y.C. Liang, Y. Maimury, A.H.L. Chen, and J.R.C. Juarez, "Machine Learning-Based Prediction of Air Quality," *Applied Sciences*, Vol. 10, Issue 24, p. 9151, 2020.

[13] U. Pak, C. Kim, U. Ryu, K. Sok, and S. Pak, "A Hybrid Model based on Convolutional Neural Networks and Long Short-term Memory for Ozone Concentration Prediction," *Air Quality, Atmosphere & Health*, Vol. 11, Issue 8, pp. 883-895, 2018.

[14] A. Chakma, B. Vizena, T. Cao, J. Lin, and J. Zhang, "Image-based Air Quality Analysis using Deep Convolutional Neural Network," *IEEE International Conference on Image Processing (ICIP)*, pp. 3949-3952, 2017.

[15] M.G. Ragab, S.J. Abdulkadir, N. Aziz, Q. Al-Tashi, Y. Alyousifi, H. Alhussian, and A. Alqushaibi, "A Novel One-dimensional CNN with Exponential Adaptive Gradients for Air Pollution Index Prediction," *Sustainability*, Vol. 12, Issue 23, p. 10090, 2020.

[16] S. Du, T. Li, Y. Yang, and S.J. Horng, "Deep Air Quality Forecasting using Hybrid Deep Learning Framework," *IEEE Transactions on Knowledge and Data Engineering*, Vol. 33, Issue 6, pp. 2412-2424, 2019.

[17] S. Srikamdee and J. Onpans, "Forecasting Daily Air Quality in Northern Thailand Using Machine Learning Techniques," *4th International Conference on Information Technology (InCIT)*, pp. 259-263, 2019.

[18] X.Y. Ni, H. Huang, and W.P. Du, "Relevance Analysis and Short-term Prediction of PM2.5 Concentrations in Beijing based on Multi-source Data," *Atmospheric Environment*, Vol. 150, pp. 146-161, 2017.

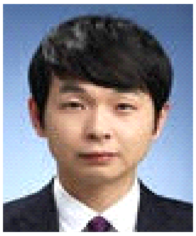
[19] N. Tomar, D. Patel, and A. Jain, "Air Quality Index Forecasting using Auto-regression Models," *IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS)*, pp. 1-5, 2020.

[20] J. Han, H. Liu, H. Zhu, H. Xiong, and D. Dou, "Joint Air Quality and Weather Prediction Based on Multi-Adversarial Spatiotemporal Networks," *Proceedings of the 35th AAAI Conference on Artificial Intelligence*, pp. 4081-4089, 2021.

[21] X. Sun, S. Li, X. Chen, and K. Wang, "Air Quality Index Prediction Based on Improved PSO-BP," *IOP Conference Series: Earth and Environmental Science*, Vol. 692, No. 3, p. 032069, 2021.

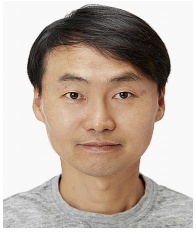
Muneeb A. Khan

2014: Computer Engineering from the Department of Electrical and Computer Engineering, COMSATS University, Pakistan (B.S.)
2019: Information Technology (IT) from the School of Electrical Engineering and Computer Science, National University of Sciences and Technology (NUST), Pakistan (M.S.)
2021~Present: Ph.D. Student, Department of Software, Sangmyung University
Research Interests: Internet of Things, artificial intelligence, computer vision, reinforcement learning and pervasive computing



Hyun-chul Kim

1995: Department of Computer Science, KAIST (B.S.)
1997: Department of Computer Science, KAIST (M.S.)
2005: Department of Computer Science, KAIST (Ph.D.)
2012~Present: Associate Professor, Department of Software, Sangmyung University
Research Interests: Data Science, Applied Data Mining, Networks



Heemin Park

1993: Department of Computer Science and Engineering, Sogang University (B.S.)
1995: Department of Computer Science and Engineering, Sogang University (M.S.)
2006: Department of Electrical and Computer Engineering, University of California, Los Angeles (Ph.D.)
2013~Present: Associate Professor, Department of Software, Sangmyung University
Research Interests: Artificial Intelligence, Computer Vision, Intelligent Systems, Internet of Things