Air Quality Analysis and PM2.5 Concentration Prediction Using Deep Learning Techniques

LITERATURE REVIEW

MOHAMED AIT AMOU- 201950000046

HEBEI UNIVERSITY OF TECHNOLOGY – 河北工业大学 |

Literature Review

Mohamed AIT AMOU, Pr. Xia Kevin

med.aitamou@outlook.com

Hebei University of technology, School of Electronic and Information Engineering

*Correspondence: 201950000046@stu.hebut.edu.cn ; Tel.: +86,156-1852-1493

Abstract: In recent years, Air Pollution has become a serious Concern in Many Cities in China due to its impact on Human health include heart disease, lung cancer, and respiratory diseases, Many researchers Paid more attention to this topic to find strategies and solutions to Reduce Air Pollution in Urban Areas, The PM2.5 '' fine inhalable particles, with diameters that are generally 2.5 micrometers and smaller '' **(1)**, is considered as one of the most dangerous pollutants that affect air quality and is responsible for respiratory diseases, which makes it a major cause of death among citizens, in This literature review we Focused on the Various Deep learning Algorithms used to Predict the PM2.5 Concentration so that the authorities can take actions to reduce it, Among this Studies many researchers used Artificial Neural Network, Long-Short-Term-Memory, Random Forest, Multilayer Perceptron, Linear Regression, Auto-regression Algorithms to Predict the Concentration of PM2.5.

Keywords: Air Pollution, PM2,5 Concentration, Prediction, Deep Learning LSTM, ANN, Random Forest (RF), Multilayer Perceptron (MLP)

I. Introduction

Since the industrial and economic revolution, the world's air pollution is growing rapidly. Most industrial activities emit huge amounts of toxic and harmful pollutants into the atmosphere such as **PM2.5** (diameter less than 2.5 micrometers), **PM10** (diameter less than 10 micrometers), **SO2** (Sulphur dioxide), **NO2** (Nitrogen dioxide), etc... Most of these air pollutants seriously affect people's health. For example, respiratory, heart and lung diseases by inhaling these pollutants, as World Health Organization estimates that around **7 million people die** every year from exposure to polluted air. And Ambient air pollution alone caused some **4.2 million deaths** in 2016 (1). According to study by (2) Particulate air pollution **(PM (2.5))** was responsible for around 3% of deaths from heart and lung disease, around 5% of deaths from cancer of the trachea, bronchi and lung disease, and about 1% of deaths from acute respiratory infections in children under five years of age worldwide. This equates to approximately 0.8 million (1.2%) premature deaths and 6.4 million (0.5%) years of life lost (YLL).

II. Overview of Air Quality

Air Pollution has become a Hot topic among data and environmental scientists who want to make a difference in the global impact of reducing air pollution. Many studies shows that's PM2.5 (diameter less than 2.5 micrometers), is the main Pollutant in Many Chinese cities and causes significant health impacts and economic losses (3) , According to Report on the state of the Ecology and Environment in China in **2016** Provided by Ministry of Ecology and Environment, the people's republic of China, in 388 big cities in china , the range of annual average concentrations of PM2.5 was **12-158 µg/m3**, with an average value of **47 µg/m3**, while the PM2.5 was the main pollutant for more than **80.3%** of the days with server pollution , in **2017**, from 338 cities, **2,311** days were under sever pollution and 802 days severely contaminated. Among them, days with PM2.5 was the primary polluted took up to **74.2%,** The average annual concentration of PM2.5 was **10-86 µg/m3**, with an average level of **43 µg/m3,** (4) (5)**.**

PM2.5 **(Figure1)** stands for particulate matter (also called particle pollution): the term for a mixture of solid particles and liquid droplets found in the air. Some particles, such as dust, dirt, soot, or smoke, are large or dark enough to be seen with the naked eye. Others are so small they can only be detected using an electron microscope, the average human hair is about 70 micrometers in diameter making it 30 times larger than the largest fine particle (6).

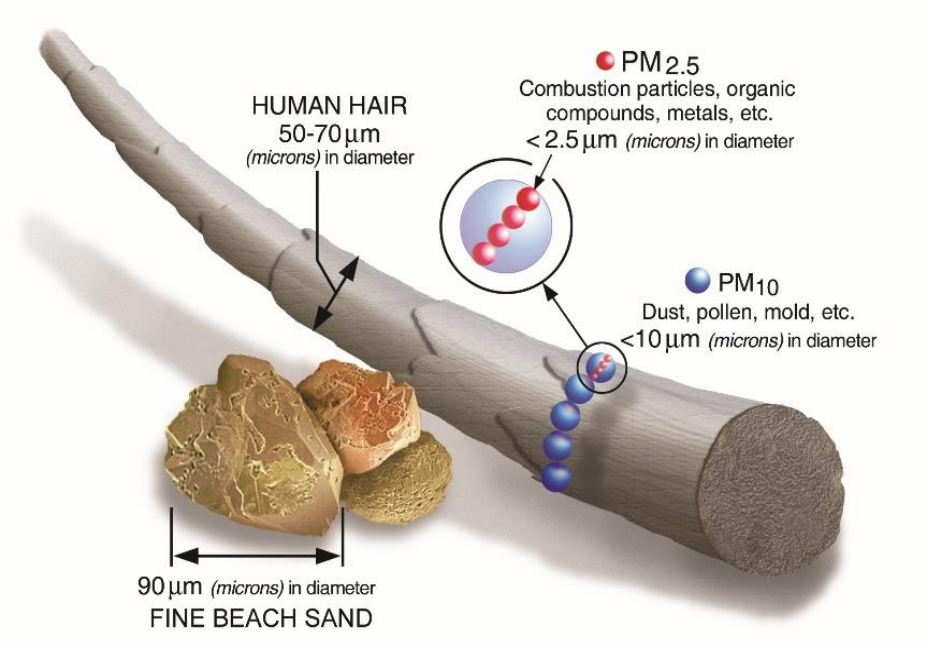

Figure 1 : Size Comparisons of PM2.5, PM10 and Human Hair *(6)*.

In 2017, exposure to PM2.5 was the third leading risk factor for type 2 diabetes deaths and DALYs, after high blood sugar and high body mass index (7), all this factors shows that Air Quality is a massive problem in worldwide especially china due to the

industrialization and Population grow in the Recent decades, Many researchers used different methods to estimate the PM2.5 concentration Recently, such as :

- Spatial Interpolation
- Remote Sensing Technique
- Air Quality Model Methods
- Machine Learning Methods

a. Spatial Interpolation:

Many Researchers used Spatial Interpolation techniques to estimate the PM2.5 Concentration, noting that This is one of best method to use to estimate the unknown data, and the most used methods are the Inverse distance weighted (**IDW**), Ordinary Kriging (**OK**) Interpolation and Trend surface Interpolation (**TS**) (8) (9), nearest neighbor interpolation (NNI), Spatial averaging Interpolation (**SAI**) (10).

b. Remote Sensing Techniques:

Air Quality monitoring is one of the most important applications of satellite remote sensing, Remote sensing formulas are used worldwide (11) (12) (13), In the recent years many researcher proposed a complex models to describe the relationship between PM2.5 and Aerosol Optical depth (AOP), (14) Satellite remote sensing data have the following unique properties:

- Near-simultaneous view over a large area
- Global coverage
- Good spatial resolution.

c. Air Quality Model Methods:

Air quality Models evaluate the relationship between air pollutant emissions and their resulting concentration in the ambient air (15). Commonly used air quality models are:

- Conceptual Models
- Emission Models
- Meteorological Models
- Chemical Models
- Source Oriented Models
- Receptor Models.

These models are important for the EPA's air quality management system because they are widely used by air pollution control agencies to identify source contributions

to air quality problems and help design effective strategies to reduce harmful air pollutants (16).

d. Machine Learning Methods:

Recently There are Serval machine learning and deep learning Methods used to predict the PM2.5 concentration, in the Section 2 we Listed most Algorithms used Approaches for predicting PM2.5 in the recent years.

this paper is organized as follows. Section 1 give an introduction about the Problematic, Section 2 describes each revised paper, including the main goal, applied methodology and obtained results. Section 3 includes a discussion based on the result of reviewing the selected papers. Finally, in section 4 we included the conclusion.

III. Literature Review

1. Multiple linear Regression:

According to a study done by **Samsuri** (2020 al.et) where they predict the PM2.5 concentration of the next three hours by using Multiple Linear Regression (MLR), the data was collected from 12 different stations around Malaysia, from the years 2005 until 2014 with the exclusion of the years 2007 to 2009, spearman correlation analysis showed that CO was strongly and positively correlated to PM10, and the best Fitted MLR Model for predicting PM10 was PM(10,t+1) first our prediction, which had a higher R-squared of 0.447 compared to PM10(t+2) (0.186) and PM10(t+3) (0.129) for PM10(t+2) and PM10(t+3) respectively (17).

K. Saithanu (et al.2014) presented a Multiple Linear regression Model to predict the PM10 concentration using Metrological data and Other pollutants concentration, the performance of regression model was measured from the validation data set with MBE= +0.5926 and S=0.6951 (18)

2. Artificial Neural Network:

A. Gholamreza et al. (2016), (19) Presented an Artificial Neural network and Markov Chain Model predict the PM2.5 Concentration in Karaj City, in this Paper authors have used Three Different Models to predict the PM2.5 concentration, these three Models are independent models, (**MLP-Multilayer propagation**) (**RBF-used Radial Bsis Function**) and Mathematical Method - Markov Chain (was used to predict the probability of occurrence of PM2.5 in different periods and to indicate air quality of the city in three forms , including good quality, sensitive and unhealthy conditions), MLP and RBF both belong the Artificial Neural network family, the RBF has an easy architecture, Model where they used only CO, NOx, PM10 and temperature as inputs

has the best performance (MLP), The results indicated that increasing the learning rate caused weakens of the network performance. On the other hand, lower values in the learning rate made the learning process time-consuming but with better performance .in term of accuracy of predicting PM2.5 Concentration RBF perform better compared to MLP.

J.B Ordieres (et al.2014) (20), Proposed a Neural network Prediction Model for PM2.5, using Multilayer Perceptron (MLP), Radial Basis Function (RBF) and Square Multilayer Perceptron (SMLP), results show that RBF take shortest training time, and combined with a greater stability during the prediction Stage, and also had best result in term of prediction of PM2.5.

Other studies **Zickus et al.** (2002) (21), **Owega et al.** (2006) (22), **Kurt et al.** (2008) (23),**Niska et al. (**2005**)** (24)**, Slini et al. (**2006**)** (25)**, Voukantsis et al. (**2011**)** (26)**, Feng and Moustris (**2013**)** (27)**,** Applied ANN models to predict air quality parameters. **Li et al. (2015)** applied ANN methods to simulate PM2.5 and PM10. Their results indicated that ANNs performed better than other methods and recommended this method as a reliable and accurate model.

3. Convolutional Neural Network:

taoying li.et (2020) (28) proposed a Hybrid **CNN-LSTM** Deep learning model to predict the PM2.5 Concentration, where she normalized the data and split it into training and testing dataset, and applied the **CNN-LSTM** proposed model to the training dataset, and evaluate the model using MAE and EMSE, the results shows that the MAE and RMSE of Hybrid **CNN**-**LSTM** Models are lower than that of **LSTM** Model, and the training time of the **CNN**-**LSTM** is also Less than the traditional **LSTM** Model

Chiou-Jye Huang (.et 2018) (29) Combined **CNN** and **LSTM** and applied it to PM2.5 Forecasting system, and compare the MAE, RMSE, Pearson correlation coefficient, IA with other Traditional machine learning methods **SVM**, **RD**,**DT**,**MLP**,**CNN** and **LSTM**, the experiment results shows that the proposed **CNN-LSTM** (Also named **APNet**) Model was best, and its MAE and RMSE were both lowest.

4. Long-Short-Term-Memory (LSTM):

Thanongsak (et.al 2020) (30) , Created two different model to predict the PM2.5 and PM10 using both LSTM and DAE algorithms using hourly measurements at the 25 monitoring stations in Seoul, the results shows that LSTM Model's Performance better than DAE, LSTM model had minimum RMSE values of 11.113 for PM10 and 12.174 for PM2.5 at a batch size of 32. At the same time, the DAEs model had minimum RMSE values of 15.038 for PM10 and 15.431 for PM2.5 at a batch size of 64, and LSTM perform better than DAE after comparing the Total average of RMSE

References

1. **World Health Organization.** ***who.net.*** **[Online] https://www.who.int/news-room/air-pollution#:~:text=Air%20pollution%20levels%20remain%20dangerously%20high%20in%20many,die%20every%20year%20from%20exposure%20to%20polluted%20air..**

2. ***The global burden of disease due to outdoor air pollution.*** **Aaron J Cohen 1, H Ross Anderson, Bart Ostro, Kiran Dev Pandey, Michal Krzyzanowski, Nino Künzli, Kersten Gutschmidt, Arden Pope, Isabelle Romieu, Jonathan M Samet, Kirk Smith. 2005 Jul 9-23.**

3. ***A review on health cost accounting of air pollution in China.*** **R. Bai, J.C.K. Lam and V.O.K. Li. 2018, Vols. pp. 279-294.**

4. 2017 Report on the State of the ecology and environment in China. [Online] 2017. http://english.mee.gov.cn/Resources/Reports/soe/SOEE2017/201808/P020180801597738742758.pdf.

5. 2016 Report on the state of the environmenet in China . [Online] 2016. http://english.mee.gov.cn/Resources/Reports/soe/ReportSOE/201709/P020170929573904364594.pdf.

6. United States Environmental Protection Agency. [Online] https://www.epa.gov/pm-pollution/particulate-matter-pm-basics.

7. State of Global Air. [Online] 2019. https://www.stateofglobalair.org/sites/default/files/soga_2019_report.pdf.

8. ***Chemical characterisation of PM2.5, PM10 and coarse particles at urban, near-city and rural sites in Switzerland.*** **Christoph Hueglin, Robert Gehrig,Urs Baltensperger,Martin Gysel,Christian Monn,Heinz Vonmont. pages 637-651, February 2005, Vols. Volume 39, Issue 4.**

9. ***High-Resolution Satellite-Derived PM2.5 from Optimal Estimation and Geographically Weighted Regression over North America.*** **van Donkelaar A, Martin RV, Spurr RJ, Burnett RT. 2015 Sep .**

10. ***Comparison of spatial interpolation methods for the estimation of air quality data. J Expo Sci Environ Epidemiol .*** **Wong, D., Yuan, L. & Perlin, S. 13 September 2004.**

11. ***Remote sensing of atmospheric particulate mass of dry PM2.5 near the ground: Method validation using ground-based measurements.*** **Li, Z., et al. 2016.**

12. ***Using satellite remote sensing data to estimate the high-resolution distribution of ground-level PM2.5.*** **Lin, C., et al. 2015.**

13. ***Remote sensing of atmospheric fine particulate matter (PM2.5) mass concentration near the ground from satellite observation.*** **Zhang, Y. and Li, Z. 2015.**

14. ***The Applicability of Remote Sensing in the Field of Air Pollution .*** **P. Veefkind+R.F. van Oss+H. Eskes+, A. Borowiak*, F. Dentner* and J. Wilson*. 2007.**

15. ***Urban Air Pollution Modeling, Air Quality - Models and Applications, Dragana Popovi.*** **Rao, Anjali Srivastava and B. Padma S. (July 5th 2011).**

16. United State Environmental Agency. ***epa.*** **[Online] https://www.epa.gov/scram/air-quality-models.**

17. ***Development of Multiple Linear Regression for Particulate Matter (PM10) Forecasting during Episodic Transboundary Haze Event in Malaysia.*** **Samsuri Abdullah 1, * , Nur Nazmi Liyana Mohd Napi 1 , Ali Najah Ahmed 2,3 , Wan Nurdiyana Wan Mansor 1 , Amalina Abu Mansor 4 , Marzuki Ismail 4,5 , Ahmad Makmom Abdullah 6 and Zamzam Tuah Ahmad Ramly. 2020.**

18. ***Using Multiple Linear Regression To Predict PM10 Concentration In Chonburi, Thailand.*** **Mekparyup2, K. Saithanu1 and J. 2014 : s.n.**

19. ***Predicting PM2.5 Concentrations Using Artificial Neural Networks and Markov Chain, a Case Study Karaj City.*** **Gholamreza Asadollahfardi*, Hossein Zangooei and Shiva Homayoun Aria. 2016.**

20. ***Neural network prediction model for fine particulate matter (PM2.5) on the USeMexico border in El Paso (Texas) and Ciudad Jua´rez (Chihuahua).*** **J.B. Ordieresa, *, E.P. Vergaraa , R.S. Capuzb , R.E. Salazarc. 2004.**

21. ***Comparison of four machine learning methods for predicting PM10 concentration in Helsinki, Finland.*** **Zickus, M., Greig, A.J., Niranjan, M. 2002.**

22. ***Identification of long-range aerosol transport patterns to Toronto via classification of back trajectories by cluster analysis and neural network techniques.*** **Owega, S., Khan, B.U.Z., Evans, G.J., Jervis, R.E., Fila,. 2006.**

23. ***An online air pollution forecasting system using neural networks.*** **Kurt, A., Gulbagci, B., Karaca, F., Alagha,. 2008.**

24. ***Evaluation of an integrated modelling system containing a multi-layer perceptron model and the numerical weather prediction model HIRLAM for the forecasting of urban airborne pollutant concentrations.*** **Niska, H., Rantamäki, M., Hiltunen, T., Karppinen, A.,Kukkonen, J., Ruuskanen,. 2008.**

25. ***PM10 forecasting for Thessaloniki, Greece.*** **Slini, T., Kaprara, A., Karatzas, K., Moussiopoulos, N. 2006.**

26. ***Intercomparison of air quality data using principal component analysis, and forecasting of PM10 and PM2.5 concentrations using artificial neural networks, in Thessaloniki and Helsinki.*** **Voukantsis, D., Karatzas, K., Kukkonen, J., Räsänen, T., Karppinen, A., Kolehmainen, M. 2011.**

27. ***Artificial neural network forecasting of PM2.5 pollution using air mass trajectory based geographic model and wavelet transformation.*** **Feng, X., Li, Q., Zhu, Y., Hou, J., Jin, L., Wang, J. 2015.**

28. ***A Hybrid CNN-LSTM Model for Forecasting.*** **TAOYING LI, MIAO HUA, AND XU WU. 2020.**

29. ***A Deep CNN-LSTM Model for Particulate Matter (PM2.5) Forecasting in Smart Cities.*** **Kuo, Chiou-Jye Huang and Ping-Huan. 2018.**

30. ***Air Pollution Prediction Using Long Short-Term Memory (LSTM) and Deep Autoencoder (DAE) Models.*** **Thanongsak Xayasouk 1, † , HwaMin Lee 2,† and Giyeol Lee. 2020.**