Global Journal of Computer Science and Technology: A Hardware & Computation
Volume 19 Issue 1 Version 1.0 Year 2019
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

A Comparative Analysis of Air Pollution Detection Technique using Image Processing, Machine Learning and Deep Learning Approach

By Samia Sultana
Stamford University Bangladesh

Abstract- Air pollution is one of the principal environmental issues for the industrial emission and infection of the atmosphere which is caused by the climatic and traffic elements, burning of fossil fuels, etc. For the past several years, various methods and models have been discovered to detect the pollution of the air. In this paper, among all of those, three mechanisms have been focused, which are image processing approach, machine learning, and deep learning technique. A comparative study has developed among these three methods to detect the pollutant of the air in the account of time, cost and efficiency so that different scenario and system can choose the best method according to their need. The objective of this paper is to assimilate the procedure of these methods in brief and utilize this study to estimate the best solution for the corresponding requirement of any particular circumstances.

Keywords: *air pollution, image processing, machine learning, deep learning.*

GJCST-A Classification: *I.2.m*

Strictly as per the compliance and regulations of:

A Comparative Analysis of Air Pollution Detection Technique using Image Processing, Machine Learning and Deep Learning Approach

Samia Sultana

Abstract- **Air pollution is one of the principal environmental issues for the industrial emission and infection of the atmosphere which is caused by the climatic and traffic elements, burning of fossil fuels, etc. For the past several years, various methods and models have been discovered to detect the pollution of the air. In this paper, among all of those, three mechanisms have been focused, which are image processing approach, machine learning, and deep learning technique. A comparative study has developed among these three methods to detect the pollutant of the air in the account of time, cost and efficiency so that different scenario and system can choose the best method according to their need. The objective of this paper is to assimilate the procedure of these methods in brief and utilize this study to estimate the best solution for the corresponding requirement of any particular circumstances.**

Keywords: air pollution, image processing, machine learning, deep learning.

I. Introduction

Air pollution is a mixture of natural and artificial substances that causes various harmful effects on human and the atmosphere. Most of the industrial activities emit massive amounts of toxic or harmful pollutants to the atmosphere such as SO2, NO2, CO, PM, and toxic organics. The pollution may also Lead to more serious problems affecting people and the whole world, such as global warming and climate change. The main reason for air quality damage is due to the smoke exhaust from industries, pollution generated by power plants, and the smoke exhaust from various vehicles. For the past few years, many methods and procedure have been invented and followed to detect air pollution. By image processing procedure, polluted images are collected from the environment and compared with the footages which are pollution-free. From those images, the diffusion process has been done, and the ratio factor is obtained to get the level of pollution. However, it works well for the images of a higher level of noise [1]. Again, if we consider the machine learning method, it detects the PM (Particulate matter) 2.5 levels based on atmospheric value for a particular day. Logistic regression is used to identify if a data sample is polluted or not, and auto-regression evaluates the future PM2.5 value based on past PM2.5 values [2]. Lastly, about the deep learning approach, which is a sub-cluster of machine learning, it uses large data set, solve the problem without dividing, using more layers, processing sequential layers simultaneously [3]. Since air pollution is a very hazardous factor for not only living being but also for nature, several numbers of techniques and procedures are there to detect this. In most of the cases, the main concentration is always on a single method and its analysis. But in this paper, the main focus will be on these three methods- pros and cons, expense and accuracy of these techniques. Comparison and correlation of these methods in one study might be convenient to differentiate, and select the suitable approach accordingly in many essential conditions.

Author: Samia Sultana, Lecturer, Department of Computer Science and Engineering, Stamford University Bangladesh.
e-mail: sayysamia@gmail.com

II. Literature Review

Air pollution is a modern-day curse, which is an outcome of expanding urbanization and industrialization. It does, however, evolve with interesting transitions in line with economic, technological, and political change [4]. So, to disseminate the impact, the detection of pollution is necessary at the very first place. Among several detection procedures, a brief background description is given below of the three focused methods of this paper-

a) Image processing

Image processing is an effective method of converting an image, perform some necessary operation upon it, and extract useful information accordingly [5]. Some common important steps of image processing are- image pre-processing, segmentation, compression, canny edge detection, grayscale conversion, Gaussian blur, edge tracking by hysteresis, and many more.

Various techniques of image processing are used to detect contamination of the air. For example, satellite images are processed, and observed on different frequencies for being compared with different surface measurement to detect the pollution [6]. In some other research, IVS camera is used to obtain the reflectance caused by atmospheric components. [7].

b) Machine Learning

Machine learning is an utilization of artificial intelligence (AI) that enable the systems to learn explicitly and enhance the output from experience [8]. The procedure of learning begins with observations, experience, and instruction to seek patterns in data and make better observation, and decisions.

Nowadays several machine learning methods have been used to predict and detect the air pollution such as- the extreme learning machine (ELM), online sequential multiple linear regression (OSMLR), etc, which have been successful in forecasting ozone and also PM2.5 [9].

c) Deep Learning Approach

Deep Learning is a subfield of machine learning concerned with algorithms inspired by the structure and function of the brain called artificial neural networks [10]. When the quantity of data intensified, machine learning techniques are insufficient in terms of performance, and deep learning gives a better performance like accuracy. Deep Learning methods can also provide productive result within air pollution epidemiology as it focuses on standard artificial neural network algorithms but operates on much large-scale, and deeper networks. The training operation with the combination of the depth of the networks allows the learning and proficiency of data abstractions at the different depths [11].

In this manuscript, out of so many procedures, only three simple techniques of these three methods will be described. The main objective of this paper is to give a concise knowledge on some significant air pollution detection procedure so that new and interested researchers in this field can get a reflection of some well-established methods in one paper and also synthesize a comparison among these methods.

III. Methodology

Considering the three procedures mentioned above, each of them performs and operates differently to detect air pollution. To estimate the better comparison, how they work and processed in the air pollution detection methodology, are discussed below-

a) Image processing

In this process, to detect the pollutants, images are obtained from the environment. The binary segmentation algorithm is used to segment the input image. Various steps such as image pre-processing, edge detection have been followed to obtain the pollution level in the environment using canny operator. Pre-processing of an image is used to improve the image quality of the input image. It partitions the images into small frames to obtain clarity regarding the frame which has the smoke or pollution content. Noise brightness ratio is used to find the ratio of the noisy image to that of the noiseless image. The diffusion process is used to remove unwanted noises (such as trees) and enhance the image quality to show only the smoky or polluted area [1]. Finally, a dialog box is displayed to show the level of pollution. With the help of input images, the System Monitor the pollution and obtain the ratio, and the diffusion process.

b) Machine Learning

The machine learning experiment has done on data set to detect the air quality was obtained from the UCI repository. The dataset have the following attributes:

- Temperature
- Wind speed
- Dew point
- Pressure
- PM2.5 Concentration(ug/m ^ 3)

Logistic regression is an algorithm employed to detect if a user-defined sample to be polluted or not. The data set gets classified into two classes —polluted or not polluted. Like all regression analyses, the logistic regression is a predictive analysis. The logit function is used to generate log odds of an attribute that signifies the probability of the attribute. Based on the logit function, the system classifies the training data to be either 0 (not polluted) or 1 (polluted) and verifies its accuracy using the test data. The result of the user input is also 0/1 and not the PM2.5 level [2].

Basically, the forecasting and prediction of the level of PM2.5 of the future condition depend on the past assessments. An autoregressive (AR) model considers observations from previous time steps as input to predict the value at the next time step. Auto-regression applied on time series data set to predict the PM2.5 rate seven days before the current date and the date on which the value of PM2.5 is to be predicted [2]. And the prediction is pretty close if we compare it to the actual and final one.

c) Deep Learning Approach

There are several modeling techniques appropriate for air pollution prediction in deep learning. LSTM (Long short term memory) method is the mostly used one for this purpose. LSTM model employs a framework to estimate future forecasting, the contamination and meteorological information of time series data and it is also a part of recurrent neural networks (RNN). In LSTM model, a memory block is used instead of neurons in hidden layer of standard RNN.

In this context, another famous method is the STDL (Spatiotemporal deep learning) which take into account spatial and temporal variations for prediction. Stacked auto-encoder models are used as an introduction model to remove inherent air quality features. The main idea behind stacked auto-encoder is the connected output layer of auto-encoder stacked in below layer is wired to the successive input layer [3].

Moreover, spatiotemporal data analysis is prominent for increasing prediction performance in DAL (deep air learning) models which mainly uses feature selection and semi-supervised learning. DAL is an efficient methodology which also considers spatiotemporal semi-supervised learning and feature selection in the input and output layers [3].

Hence, all these deep learning methods give a reasonably good prediction of various pollutants of the air like- PM2.5, O3, NO2, and PM10.

IV. Comparison

If we compare among these three, all of them can find about the pollution to some extent. But if we follow each of them one by one, in image processing, it only identifies if the air is pollutant or not. It gives output just in a dialog box and defines a kind of yes or no. It does not exaggerate any further. So, it is only able to detect the pollution if only the above particular image processing technique is concerned. Again, if we observe the machine learning method, it works on a specific data set, first find out if the air is pollutant or not. Then, it tries to find out the level of PM2.5. So, in comparison with image processing, after detecting the pollution, it is little more specified since it finds the PM2.5 level as well considering the above method of machine learning.

Lastly, looking at the deep learning approach and its various methods, it finds out not only PM2.5 but also some other pollutant along with the detection of pollution. But it works on massive data set in comparison with machine learning.

For better understanding, we can observe the following table which provides fundamental differences and similarities of the references [1] [2] [3].

Table 1: Comparison of the three methods

	Image Processing	Machine Learning	Deep Learning
Pollution Detection	Possible.	Possible.	Possible.
Dataset	Not needed.	Needed.	Needed.
Accuracy Level	Good, considering the detection.	Very good, As identify PM2.5 level along with detection.	Excellent, As identify PM2.5, O3, NO2, and PM10 level along with detection.
Pollutant Detection	None.	PM2.5.	PM2.5, O3, NO2, and PM10.
Expense	Less expensive regarding no need of data set.	Expensive as requirement of dataset.	Most expensive considering huge range of dataset.
Detection area	Particular area where the camera is placed and can move.	Large area considering the range of data set.	Large area considering the range of data set.

Therefore if we observe the three procedures that have been discussed in this paper, considering the effectiveness and accuracy, deep learning approach will give the best result among these three. But it needs to be done on a huge data set. So along with its accuracy level, the overall expense and cost will also increase. If simple detection is the requirement, then image processing can be a good option.

V. Conclusion

Air Pollution is one of the major issues of nature, and it is expanding day by day with increasing urbanization and industrialization. Several methodologies are there to detect that. However, in this manuscript, as mentioned earlier, the focus was on only three simple procedures. Hence, the primary observations of this paper are- image processing can be a favorable choice for detection, but more specified algorithm and sensors are needed to detect the impure substance. But if more specified detection of pollutants is required, then machine learning or deep learning will be a better approach. And if the expense is concerned, deep learning will be most expensive owing to the enormous range of the dataset. In the future, more enhanced mechanism may discover in regard to both cost and accuracy.

References Références Referencias

1. Joans. S. M and Pavithra. M, (2017) Air Pollution Monitoring through Image Processing, IJRASET, 5, 1377-1382. https://www.ijraset.com/fileserve.php?FID=10557
2. C R. A, Deshmukh C. R., D K. N and Vidyavastu P. G (2018) Detection and Prediction of Air pollution using Machin Learning Models, IJETT, 59, 204-207. http://ijettjournal.org/2018/volume-59/number-4/IJETT-V59P238.pdf
3. Ayturan Y. A, Ayturan Z. C and Altun H. O (2017) Air Pollution Modeling with Deep Learning: A Review, IJEPEM, 1(3),58-62. https://dergipark.org.tr/download/article-file/555329
4. Nathanson J. A (2018) Air Pollution https://www.britannica.com/science/air-pollution

5. Elsevier B. V (2019) Image Processing https://www.sciencedirect.com/topics/neuroscience/image-processing
6. Prochdxka. A, Kolinovd. M, Fiala. J, Hampl. P and Hlavaty. k (2000), Satellite image processing and air pollution detection, 2000 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No.00CH37100), Istanbul, Turkey, 2000, pp. 2282-2285 vol.4.
7. By C. J. Wong, MatJafri. M. Z., Abdullah. K and. Lim H. S (2009), Development of Low cost Air quality Monitoring System using Image Processing Technique. Yung-Sheng Chen, Intech Open.
8. Fraggella.D (2019) What is Machine Learning https://emerj.com/ai-glossary-terms/what-is-machine-learning/
9. Bellinger, C., Mohomed Jabbar, M. S., Zaïane, O., & Osornio-Vargas, A. (2017). A systematic review of data mining and machine learning for air pollution epidemiology. BMC public health, 17(1), 907.
10. Brownlee. J (2016) Deep Learning and Artificial NeuralNetwork.https://machinelearningmastery.com/what-is-deep-learning/
11. Peng. H (2013), Air Quality Prediction by Machine Learning Methods, B.Sc. Thesis, Sun Yat-sen University, Vancouver.