Study on Effect of Air Pollution on Health

Isha Aggarwal, Pranshu Tyagi
Student
Bhagwan Parshuram Institute of Technology

Abstract - The main objective of the air quality index (AQI) system is to interpret air quality in a standard indicator to enable public understand the likely health and environmental impacts of air pollutant concentration levels monitored on any given day. The enormity of air pollution has always been a matter of concern due to rapid development and urbanization over a long period. In our previous studies conducted, people have analyzed the data till 2010. Detailed analysis from 2003-2015 of air pollutants has been proposed in this extended paper. Correlation analysis have been done to study the relation between various air pollutants like sulphur dioxide (SO_2), nitrogen dioxide (NO_2), suspended particulate matter (PM) and various health effects on human.

Index Terms - Air quality index, health impacts, mortality, Correlation analysis

---

I. INTRODUCTION

According to World Population Review, Delhi, the National Capital Territory (NCT) of India, is the densely populated metropolitan city with a large influx of population from other states of India. As per the last Census carried out in 2011, population of Delhi was 16.7 million [1] and estimated 2016 population of 18.6 million. In recent years, rapid industrialization and urbanisation posed detrimental effect on environment. Problem of air pollution is increasingly getting more serious. Increasing levels of pollutants in air is causing extreme health disorder. It directly affects a population of millions who are suffering from shortness of breath, eye irritation to chronic respiratory disorders, pneumonia, acute asthma etc [2] [3].
Increasing levels of pollutants in air is causing extreme health disorder. It directly affects a population of millions who are suffering from shortness of breath, eye irritation to chronic respiratory disorders, pneumonia, acute asthma etc [4]. Watery eyes, coughing and difficulty breathing are acute and common reactions of air pollution. Air pollution has both acute and chronic effects on human health, affecting a number of different systems and organs. It ranges from minor upper respiratory irritation to chronic respiratory and heart disease, lung cancer, acute respiratory infections in children and chronic bronchitis in adults, aggravating pre-existing heart and lung disease, or asthmatic attacks. In addition, short- and long-term exposures have also been linked with premature mortality and reduced life expectancy [5].
Many studies have observed associations between particulate air pollution and human health [17]. Increases in sickness and death associated with severe air pollution episodes have been well documented. Particulate air pollution is a mixture of particles that vary in size, composition, and origin. Fine particles are the largest health concern because they can be breathed most deeply into the lung. This size range includes most sulfate particles. Unlike larger particles which are derived primarily from soil and other crustal materials, fine particles are derived chiefly from combustion of fossil fuels in processes such as transportation, manufacturing, and power generation. Sulfate particles are commonly generated by conversion from primary sulfur emissions and a varying portion of sulfate particles may be acidic.

Type of pollutants

The major concentration of pollutants in the Delhi air is:-

1. Particulate Matter, RSPM and SPM (PM2.5 and PM10): The principle source of particulate matter in Delhi is vehicular emissions, particularly from heavy motor diesel vehicle, kerb-side dust, thermal power plants, industrial and residential combustion processes. Respirable suspended particulate matter (PM2.5) is considered to be more hazardous to human health than PM10. The average limit of PM2.5 pollution is 60 microgram per cubic meter but all the areas of Delhi have the level of PM 2.5 exceeding 300 microgram per cubic meter [7].
2. Nitrogen Oxide (NO_x): Oxides of Nitrogen are produced during industrial combustion processes and primarily as vehicular exhaust. NO_x levels are highest in urban areas as it is related to traffic. It is an important ingredient in generation of photochemical smog which envelops the urban air with haze like blanket. It has harmful effects such as wide-range of respiratory problems in adults and children.
3. Sulphur Dioxide (SO_2): It is formed mostly by burning of fossil fuels particularly from thermal power plant. This pollutant is the reason for acid rain and has adverse effects on lung functions.
4. Benzene: The main sources of benzene are from vehicle exhaust and other industrial processes since it is an industrial solvent. Benzene is a component of crude oil and petrol. Apart from vehicle exhaust, evaporation from petrol filling stations can raise benzene levels [7].
5. Ozone (O_3): Formed by chemical reaction of volatile organic compounds and nitrogen dioxide in the presence of sunlight, so level of ozone is generally higher in the summer. Ground level ozone also contributes in formation of photochemical smog.
6. Toluene**:** Toluene is another industrial volatile solvent whose short term exposure causes irritation of eyes and the respiratory tract. The substance is a known carcinogen and affects the central nervous system also.

7. Carbon Monoxide (CO): CO is a toxic air pollutant which is produced by incomplete combustion of carbon-containing fuels. Vehicle deceleration and idling vehicle engines are one of its main causes.

Effects on Health

- Sulfur dioxide(SO_2) is also considered to cause cough shortness of breath, spasm of the larynx and acute irritation to the membranes of the eyes. It also acts an allergic agent. When it reacts with some compound, sulfuric acid is formed which may damage lungs.
- Carbon Monoxide(CO) often affects the oxygen carrying capacity of blood. Nitric oxide is reported to be a pulmonary irritant and excess concentration may cause pulmonary hemorrhage.
- Hydrogen Sulfide is also toxic. Lead emitted from automobile exhausts is a cumulative poison and is dangerous particularly to children and may cause brain damage.
- The particulate pollutants are capable of exerting a noxious local action in the interstitial areas of the lungs. Radioactive elements are also harmful to man and other living organisms.
- Smog also has killer effects which is also the result of air pollution.

Factors leading to pollution

- Burning of rice stubble left after harvesting by farmers in Punjab, Haryana and Western Uttar Pradesh in the month of October.
- Dust pollution by construction activities (metro, buildings etc.) in the NCR region.
- Pollution of vehicles which is rapidly increasing every day adding to the reason of this invisibility
- Industrial pollution by factories or small scale industries.
- Bursting firecrackers on Diwali made the situation worse. Also, the ban on firecrackers in Delhi NCR region a few days before Diwali this year did not affect the burning of firecrackers.

II. RELATED WORK

Considerable research has been carried out in the subject area of air pollution and health data analysis along with its interpretation. Researchers Maureen L. Cropper, Nathalie B. Simon, Anna Alberini, and P.K. Sharma, have compared the impact of particulate matter of India and US and also death by age groups of India and US [7]. This paper reports the results of a study relating levels of particulate matter to daily deaths in Delhi, India between 1991 and 1994. Other researchers working in the field of air quality and health effects studies are Sanjoy Maji, Sirajuddin Ahmed and Weqar Ahmad Siddiqui, have studied daily averaged concentration data of air pollutants of monitoring sites under the National Ambient Air Quality Monitoring Programme of Delhi were analysed for the period 2001–2010 (10 years) using the AQI system [8]. Keiko Hirota , Shogo Sakamoto , Satoshi Shibuya and Shigeru Kashima discussed methods of estimating the health effects of air pollution in large Asian cities [9]. Thereafter C. Arden Pope, Michael J. Thun, Mohan M. Namboodiri, Douglas W. Dockery, John S. Evans, Frak E. Speizer and Clark W. Heath linked ambient air pollution data from 151 U.S. metropolitan areas in 1980 with individual risk factor [10]. This paper discussed methods of estimating the health effects of air pollution in large Asian cities. Akindayo Olanrewaju Ogunbayo studied the relationship between exposure to environmental air pollution and hypertensive disease[11]. A comprehensive emission inventory for Delhi, India, for 1990–2000(10 years) has been developed in support of air quality by B.R. Gurjar, J.A. van Aardenne, J. Lelieveld, M. Mohan[12]. It shows that SO2 and total suspended particles (TSP) are largely emitted by thermal power plants, while the transport sector contributes most to NOx, CO and non-methane volatile organic compound (NMVOC) emissions. Ping-Wei Soh, Kai-Hsiang Chen, Jen-Wei Huang and Hone-Jay Chu explored the spatial-temporal patterns of particulate matter (PM) in Taiwan[13]. This study explores the spatial-temporal patterns of particulate matter (PM) in Taiwan.

Long term observation of mass concentrations of PM10, PM2.5, and other gaseous pollutants in Wuhan, China, including sulphur dioxide (SO2), nitrogen oxide (NOx), from 2011 to 2014 was studued by Xin Ma, Wei Gong and Zhongmin Zhu [18]. Authors demonstrated the linear relationship between PM10 and PM2.5, and reveal that PM2.5 serves as the primary pollutant in Wuhan region. According to Report on the state of health of Delhi [19], in the monsoon of 2015, Delhi was caught unawares by an outbreak of dengue. The spurt in dengue cases had brought the focus on the capital city’s failing health infrastructure.

However there is a scope for further studies due to susceptibility of population towards increasing air pollution and to find ways to monitor and control it in effective manner.

In the presented work, correlation analysis for analyzing the trends of air pollution with health in Delhi has been used.

III. PROPOSED WORK

Proposed approach

A systematic approach has been followed in this analysis which is depicted in figure 1. The approach starts with the collection of dataset from Kaggle [15]. Collected data has been pre processed to remove the redundancy. Pre processing of data includes steps like parsing of years, cleaning and training. Further, Correlation analysis has been carried out using python libraries and then the data is analyzed.

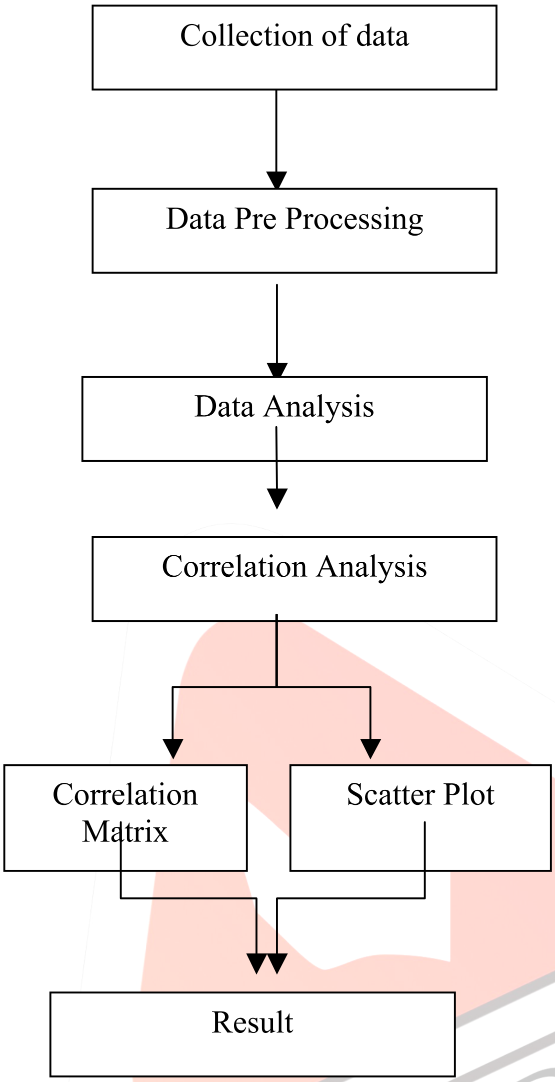

Fig 1: Flowchart of proposed approach

Data set used

A snapshot of the dataset used figure 2 contains six attributes: Year, SO_2, NO_2, RSPM, SPM and PM2.5. The 'Year' attribute describes the sampling year and other parameters give their individual centration in air. The data has been collected from 2003 to 2015.

	A	B	C	D	E	F	G
1	YEAR	SO2	NO2	RSPM	SPM	PM25	
2	2003	NA	34	149	205	NA	
3	2003	11	39	184	405	NA	
4	2003	12	42	190	429	NA	
5	2003	14	41	187	304	NA	
6	2003	13	39	192	323	NA	
7	2003	12	36	170	325	NA	
8	2003	12	43	112	262	NA	
9	2003	13	37	179	262	NA	
10	2003	11	42	155	192	NA	
11	2003	11	51	NA	245	NA	
12	2003	11	42	217	337	NA	
13	2003	9	38	71	205	NA	
14	2003	13	43	201	275	NA	
15	2003	12	44	177	302	NA	
16	2003	11	47	91	173	NA	
17	2003	14	43	102	190	NA	
18	2003	11	38	170	264	NA	
19	2003	12	42	143	243	NA	
20	2003	15	43	115	282	NA	
21	2003	10	41	119	298	NA	
22	2003	14	46	132	276	NA	
23	2003	16	44	84	266	NA	
24	2003	11	39	97	301	NA	
25	2003	15	52	131	357	NA	

Fig.2. Pollution Dataset

A snapshot of the dataset used figure 3 contains five attributes: Year, Cause, Gender, Age group and mortality. The 'Year' attribute describes the sampling year and other parameters give their individual information. The data has been collected from 2012 to 2015.

YEAR	CAUSE	GENDER	AGEGROU	MORTALITY
2012	Acute bronchitis	F	A	8
2012	Acute bronchitis	F	B	3
2012	Acute bronchitis	F	C	0
2012	Acute bronchitis	M	A	8
2012	Acute bronchitis	M	B	1
2012	Acute bronchitis	M	C	15
2012	Acute bronchitis	O	A	0
2012	Acute bronchitis	O	B	0
2012	Acute bronchitis	O	C	0
2012	Pneumonia	F	A	146
2012	Pneumonia	F	B	47
2012	Pneumonia	F	C	26
2012	Pneumonia	M	A	207
2012	Pneumonia	M	B	57
2012	Pneumonia	M	C	34
2012	Pneumonia	O	A	0
2012	Pneumonia	O	B	0
2012	Pneumonia	O	C	0
2012	Other lower respiratory disorders	F	A	2
2012	Other lower respiratory disorders	F	B	1
2012	Other lower respiratory disorders	F	C	3
2012	Other lower respiratory disorders	M	A	4
2012	Other lower respiratory disorders	M	B	3
2012	Other lower respiratory disorders	M	C	0

Fig.3. Mortality due to various diseases

A snapshot of the dataset used figure 4 contains five attributes: Year, Cause, Gender, Age group and mortality. The 'Year' attribute describes the sampling year and other parameters give their individual information. The data has been collected from 2004 to 2015

YEAR	CAUSE	GENDER	AGEGROU	MORTALITY
2004	Respiratory	M	A	243
2004	Respiratory	M	B	151
2004	Respiratory	M	C	99
2004	Respiratory	F	A	119
2004	Respiratory	F	B	119
2004	Respiratory	F	C	99
2004	Circulatory	M	A	140
2004	Circulatory	M	B	270
2004	Circulatory	M	C	244
2004	Circulatory	F	A	70
2004	Circulatory	F	B	153
2004	Circulatory	F	C	181
2005	Respiratory	M	A	8
2005	Respiratory	M	B	252
2005	Respiratory	M	C	183
2005	Respiratory	F	A	5
2005	Respiratory	F	B	184
2005	Respiratory	F	C	111
2005	Circulatory	M	A	16
2005	Circulatory	M	B	182
2005	Circulatory	M	C	184
2005	Circulatory	F	A	10
2005	Circulatory	F	B	109
2005	Circulatory	F	C	103

Fig.4. Mortality due to Circulatory and respiratory diseases

A snapshot of the dataset used figure 5 contains five attributes: Year, Cause, Gender, Age group and mortality. The 'Year' attribute describes the sampling year and other parameters give their individual information. The data has been collected from 2004 to 2015

	A	B	C	D	E	F
22	2015	probability of dying	F	C	0.016	
23	2015	probability of dying	F	D	0.051	
24	2015	probability of dying	F	E	0.159	
25	2015	probability of dying	F	F	2.028	
26	2015	expectation of life	M	A	200.8	
27	2015	expectation of life	M	B	115.9	
28	2015	expectation of life	M	C	97	
29	2015	expectation of life	M	D	140.5	
30	2015	expectation of life	M	E	62.6	
31	2015	expectation of life	M	F	46.4	
32	2015	expectation of life	F	A	210.4	
33	2015	expectation of life	F	B	123	
34	2015	expectation of life	F	C	104.2	
35	2015	expectation of life	F	D	153.2	
36	2015	expectation of life	F	E	68.4	
37	2015	expectation of life	F	F	49.8	
38	2014	age-specific death rate	M	A	0.041	
39	2014	age-specific death rate	M	B	0.002	
40	2014	age-specific death rate	M	C	0.004	
41	2014	age-specific death rate	M	D	0.019	
42	2014	age-specific death rate	M	E	0.048	
43	2014	age-specific death rate	M	F	0.467	
44	2014	age-specific death rate	F	A	0.047	
45	2014	age-specific death rate	F	B	0.002	
46	2014	age-specific death rate	F	C	0.004	

Fig 5. Mortality due to age specific, probability of dying and expectation of life

IV. METHODS AND TECHNIQUES INVOLVED

Correlation analysis

Correlation analysis has been used in the study to correlate the basic characteristics of data. We have used Pearson-type correlation. The Pearson product-moment correlation coefficient (or Pearson correlation coefficient) is a measure of the strength of a linear association between two variables and is denoted by r. The Pearson correlation coefficient, r, can take a range of values from +1 to -1. A value of 0 indicates that there is no association between the two variables. A value greater than 0 indicates a positive association; that is, as the value of one variable increases, so does the value of the other variable. A value less than 0 indicates a negative association; that is, as the value of one variable increases, the value of the other variable decreases.

Classification analysis

It is a supervised learning approach in which the computer program learns from the data input given to it and then uses this learning to classify new observation. We have used KneighborsClassifier.The k-nearest neighbour algorithm (k-NN) is a non parametric method used for classification and regression. In both cases, the input consists of the k closest training examples in the feature space. The output depends on whether k-NN is used for classification or regression. Both for classification and regression, a useful technique can be used to assign weight to the contributions of the neighbors, so that the nearer neighbors contribute more to the average than the more distant ones. The neighbors are taken from a set of objects for which the class or the object property value is known.

V. RESULTS

Results for mortality and pollution over various stations over Delhi are presented.
Correlation analysis on various diseases and circulatory & respiratory showed that there is a positive relationship of pollution and mortality. Also there is a positive linear relationship between pollution and probability of dying, expectation of life and age specific death rate for age group from 0 to 85+ years.
Scatter plot shows that with increase in pollutants over years, mortality also increases.

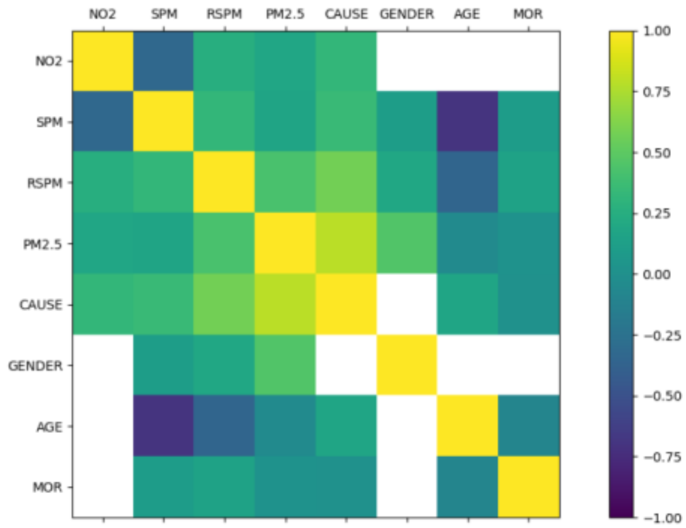

Fig 6: Correlation Matrix of various diseases vs. pollution

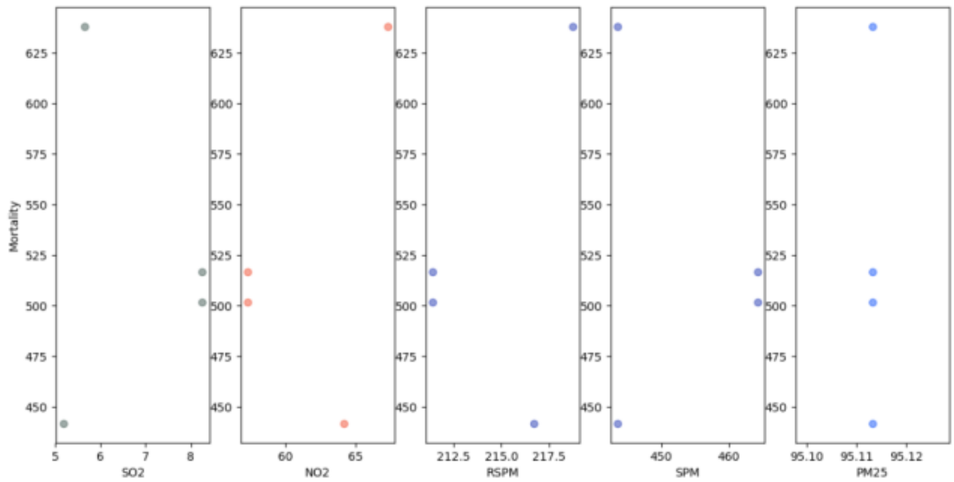

Fig 7: Scatter Plot on Pneumonia

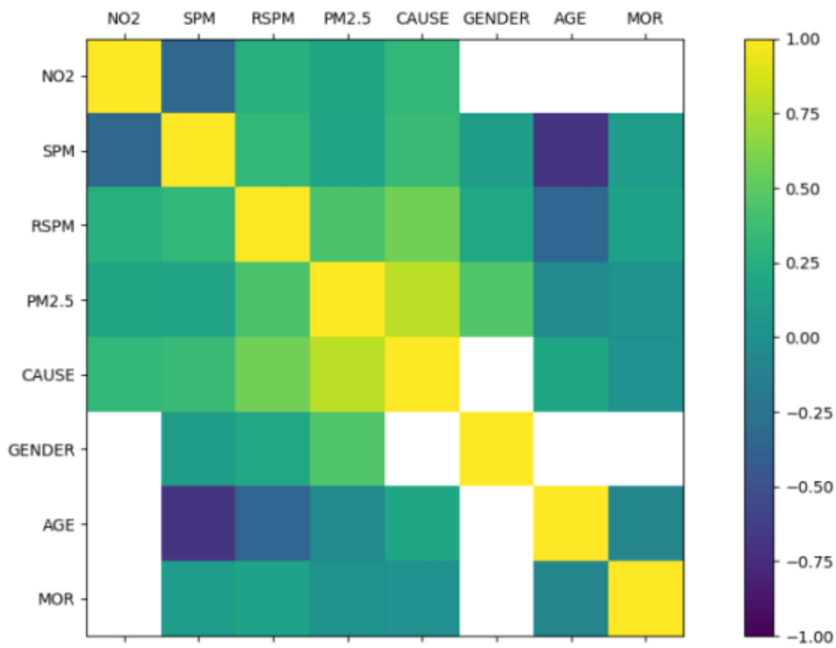

Fig. 8: Correlation Matrix of Circulatory & Respiratory vs. pollution

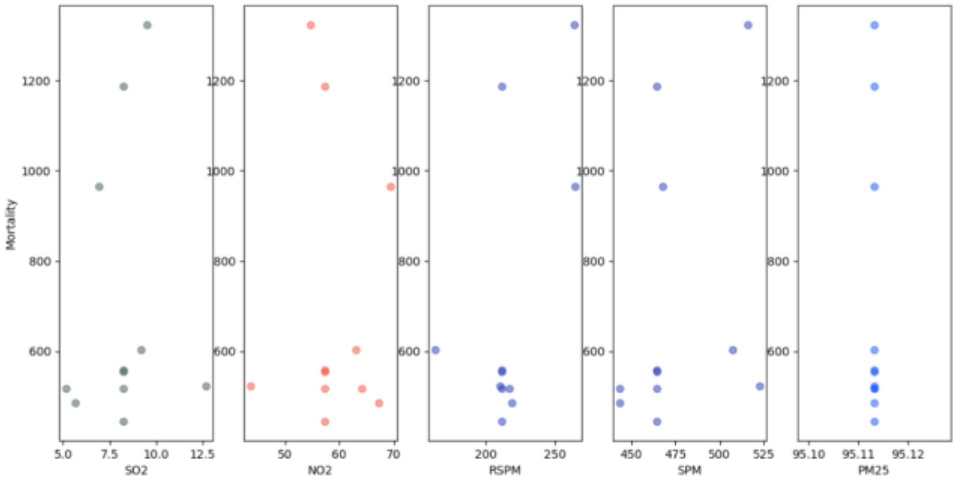

Fig. 9: Scatter Plot on Circulatory diseases

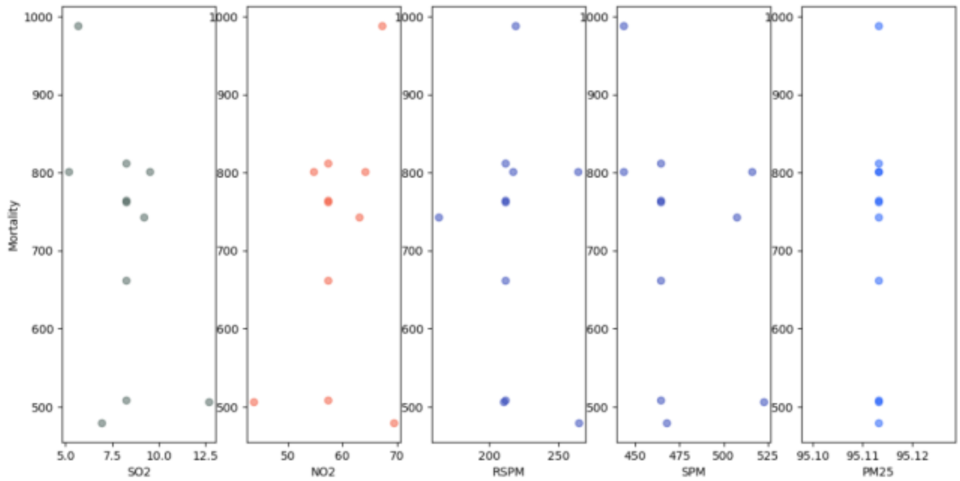

Fig. 10: Scatter Plot on Respiratory diseases

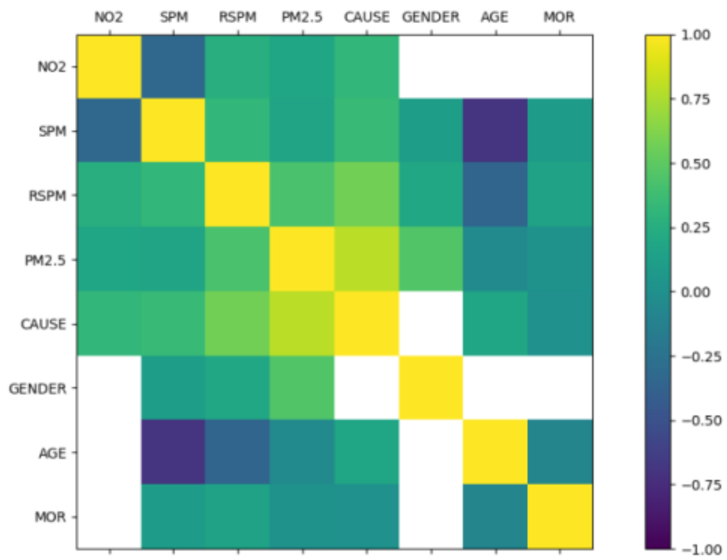

Fig. 11: Correlation Matrix of Age specific, Probability of dying and Expectation of life vs. pollution

```
-------------------------------------------------------------------------------
                    Test Accuracy Score with Summer Data
-------------------------------------------------------------------------------

Accuracy With Summer Data :  1.0

-------------------------------------------------------------------------------
                    Output Prediction with Summer Data
-------------------------------------------------------------------------------

Predicting Output for Unseen Data as used from Summer dataset [20.26,312.21,137.96,98.66,549.22
,32.09,441.14,988.29,12.13,30.74] :  ['Summer']
```

Fig. 12: Prediction of season (summer) based on test values given

```
-------------------------------------------------------------------------------
                    Test Accuracy Score with Winter Data
-------------------------------------------------------------------------------

Accuracy With Winter Data :  1.0

-------------------------------------------------------------------------------
                    Output Prediction with Winter Data
-------------------------------------------------------------------------------

Predicting Output for Unseen Data as used from winter dataset [20.26,312.21,137.96,98.66,549.22
,32.09,441.14,988.29,12.13,30.74] :  ['Winter']
```

Fig. 13: Prediction of season (winter) based on test values given

VI. Conclusion and Future Scope

The agenda of this study is to use technology for creating awareness to decrease pollution by adopting proper measures. New Delhi, which has already been ranked among the world's most polluted city, is considered in the study. This study focused on the implementation of correlation analysis, data mining techniques to fathom the different patterns in various types of pollutants. Python Libraries have been used for analyzing using Python language.
As the results indicate that the increase in pollution may attribute to increase in mortality rate in case of infants. Also it can be seen that with increase in pollution, the probability of dying in case of age group of 0-85+ years also increases.
The future work will be to explore about morbidity for all age groups. Also exploring the mortality rate in case of old age people will also be future work. In this report we have studied the morality of infants.
Pollution prevention requires public-policymakers participation which sometimes becomes a major limitation. Atmospheric variables like wind, precipitation, humidity do not remain static which also pose a problem in long range forecasting.
The results proposed are based on the historic data, hence cannot take account the future events that can manipulate the proposed results. The predicted results can be verified further from different techniques.

VII. ACKNOWLEDGEMENT

This is to acknowledge that under the guidance of the faculties I have successfully completed the case study which focuses on the Delhi pollution. This case study includes research paper on effect of air pollution on health.
However, it would not have been possible without the kind support and help of many individuals. I would like to extend my sincere thanks to all of them. I am highly indebted to Dr. Shweta Taneja and Dr. Nidhi Sharma for their guidance and constant supervision as well as for providing necessary information regarding the project & also for their support in completing the project. I would like to express my special gratitude and thanks to faculty for giving me such attention and time.

REFERENCES

[1] Statistical Abstract (2016) Delhi Govt Portal, www.delhi.gov.in.
[2] Cohen, Aaron J, Brauer, Michael et al. (2017) "Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the Global Burden of Diseases Study", *The Lancet***389 (10082)**: 1907-1908.
[3] Rizwan SA, Nongkynrih B, Gupta SK. (2013) "Air pollution in Delhi: Its Magnitude and effects on health." *Indian J Community Med***38 (1):**4-8.
[4] Dr. Nidhi Sharma, Shweta Taneja ,Vaishali Sagar, Arshita Bhatt "Forecasting air pollution load in Delhi using data analysis tools", Procedia Computer Science 00 July 2018

[5] Marilena Kampa, Elias Castanas "Human health effects of air pollution" , Environmental Pollution Volume 151, Issue 2, January 2008, Pages 362-367
[6]Realtime ambient air quality data of delhi, DPCC link : http://dpccairdata.com/dpccairdata/display.
[7] Maureen L. Cropper, Nathalie B. Simon, Anna Alberini, and P.K. Sharma "The Health Effects of Air Pollution in Delhi, India"
[8] Sanjoy Maji, Sirajuddin Ahmed and Weqar Ahmad Siddiqui "Air quality assessment and its relation to potential health impacts in Delhi, India" CURRENT SCIENCE, VOL. 109, NO. 5, 10 SEPTEMBER 2015.
[9]Keiko Hirota, Shogo Sakamoto, Satoshi Shibuya and Shigeru Kashima, "A Methodology of Health Effects Estimation from Air Pollution in Large Asian Cities" Environments 2017, 4, 60; doi:10.3390/environments4030060
[10] C.Arden Pope, Michael J. Thun, Mohan M. Namboodiri, Douglas E. Dockery, John S. Evans, Frank E. Speizer and Clark W. Heath "Particulate Air Pollution as a Predirtor of Mortality in a Prospective Study of u.S. Adults". Am J Respir Crit Care Med Vol 151 pp:669-674,1995
[11] Akindato Olanrewaju Ogubayo " Retrospective Study of Effects of Air Pollution in Human Health." , Bachelor's Thesis,Degree Programme in Environmental Engineering (Environmental Management),May 2016
[12] B.R. Gurjar, J.A. van Aardenne, J. Lelieveld, M. Mohan, "Emission estimates and trends (1990–2000) for megacity Delhi and implications" Atmospheric Environment, 28 May 2004
[13] Ping-Wei Soh, Kai-Hsiang Chen, Jen-Wei Huang and Hone-Jay Chu, "Spatial-temporal Pattern Analysis and Prediction of Air Quality in Taiwan", 2017 10th International Conference on Ubi-media Computing and Workshops (Ubi-Media)
[14] Python (2017) [Online] Available at: https://www.python.org/
[15] Data for pollution available from link: https://www.kaggle.com/
[16] Data for pollution available from link: https://www.cpcb.com/
[17] Lipfert, F. W. 1994. Air Pollution and Community Health: A Critical Review and Data Sourcebook. Van Nostrand Reinhold, New York.
[18] Xin Ma, Wei Gong, Zhongmin Zhu,"The study of long term air pollution characteristics in Wuhan,China."
[19] Report on "The State of Health of Delhi". Praha.org, Making democracy Work. July 2017