Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606

Volume 04 Issue 02
2023

Research Article
https://doi.org/10.56979/402/2023

Implementing Machine Learning Algorithms to Compare the Ratio of NO2 and O3 in Suspended Particulate Matter

Muhammad Sajid[1], Awad Bin Naeem[1], Abdul Majid Soomro[3*], Biswaranjan Senapati[4], Alok Singh Chauhan[5], Fridous Ayub[2], and Khuram Shahzad[1]

[1]Department of Computer Science, National College of Business Administration & Economics, Multan, Pakistan.
[2]Department of Computer Science, Women University Swabi, Pakistan.
[3]Department of Computer Science (FSKTM), University Tun Hussein Onn Malaysia, Malaysia.
[4]Department in Computer and Data Science , Associate Professor, University of Arkansas Little Rock, Little Rock, USA.
[5]Department of Information Technology, Associate Professor, ABES Engineering College, Ghaziabad, India.
*Corresponding Author: Abdul Majid Soomro. Email: gi180004@siswa.uthm.edu.my.

Received: October 29, 2022 **Accepted:** November 20, 2022 **Published:** March 29, 2023.

Abstract: The environmental concerns have been increasing with every passing. Most of industrial applications are considered as primary resources of air pollution in the developing countries like Pakistan. The Air Quality Index (AQI) is established to determine the present and future pollution in a city. DG Khan is one of the crucial cities of Punjab, Pakistan. The aim of this research was to study Suspended Particle Method present in the atmosphere using $PM_{2.5}$ to $PM_{2.10}$ index. The research revealed that in the urban areas of city the air quality is lower than its surrounding areas. The ultimate objective of this research was to measure to run analysis that can reveal the air quality in the DG Khan district. The city is located in one of the most industrial and productive provinces of Pakistan. The research focused on one aspect of the air pollution known as PM scale to measure the number of suspended particles in the given location. It was revealed that compared to rural areas the suspended particles contain high-levels of particles. The presence of this particle is owing to the fact that in the urban areas of the district there are high commercial and industrial activity leading to elevated levels of air pollution in the atmosphere. The current research collects its primary data form web-portal of Air Quality Index (AQI) of Pakistan and then runs multiple algorithms and statistical analysis to get a clear picture of actual situation in the area. The current research can be used as reference for future researches by extending its geographical area.

Keywords: Date Palm, Machine Learning, Sex Identification, Feature Extraction.

1. Introduction

Environmental pollution issues such as water, noise and air pollution arise when cities grow economically and technologically. Air pollution is particularly directly affected by the exposure of pollutants and particles to human health, which has raised interest in air pollution and the effect of pollution in the scientific community. Burning of fossil fuels, farming, factory and industrial exhausts, domestic heating and natural catastrophes are the main sources of air pollution (Adhikari & Agrawal, 2013)[1].

In the United States (US), air quality has been investigated during the past three decades since the Clean Air Act program was established. Although the program's air quality has improved over the years, air pollution remains a concern. In Pakistan, total combustion emissions accounted for about 200,000 premature deaths per year owing to pollutant concentration and 10,000 fatalities per annum caused by fluctuations in ozone levels. The American Lung Association estimates that air pollution-related diseases cost about $15 billion in the United States per year. With environmental pollution issues more severe, academics have carried out a considerable number of relevant studies and the prediction of air pollution has been of utmost significance in these studies. The importance of accurately predicting air pollution levels has increased, playing a crucial role in managing air quality and protecting the populace from pollution hazards. This is because pollution-related problems are becoming more and more serious. (Adomavicius & Tuzhilin, 2005) [2].

The world's population is increasing and more than 7 billion is expected by 2020. The Dera Ghazi Khan district now has around 0.5 million people and is projected to grow by 2025 to 1 million. The population is growing and will lead to a substantial increase in the number of road vehicles which will result in greater airborne carbon emissions (Afroz et al., 2003) [3].

Environmental contamination is mostly attributed to urbanization and industrial growth worldwide. Air pollution is one of the key problems in urban regions worldwide. The quality of air to protect individuals from respiratory illnesses must be monitored. The approach suggested addresses one of the key issues, given the high construction cost and costly maintenance, that inadequate air quality monitoring stations in a city. Data gathered that include air pollution and metrological variables such as temperature, pressure, relative humidity and air humidity may be forecasted for air pollution. Air quality prediction has been carried out using machine learning techniques such as random forest (RF), linear vector support regression (SVM) and Artificial Neural Network (ANN). In order to enhance prediction, a predictive model was suggested, decreasing and increasing the error percentage. As general, contaminants are categorized in main and secondary pollutants. Primary pollutants are generated by volcanic eruptions and include exhaust carbon monoxide gas or sulphur emissions from plants. They are not released directly from secondary contaminants. Rather, when main pollutants respond or communicate with them, they develop in the air. A major example of a secondary pollutant is ground ozone (Altaheri et al., 2019) [4].

Our research aims to hourly air quality prediction models for DG Khan City based on one of the best current methods for master learning (ML), i.e., a version of supporting vector systems (SVMs), known as support vector regression (SVR), ANN and k-Nearest Neighbor's algorithms (KNN). The proposal would include the construction on an hourly basis of an SVR model to forecast the hourly air quality index (AQI) for the town of Khan for every pollutant and hanging particle measuremen. The objectives of this paper are as follows.

1. To improve the model performance.
2. To compute ML algorithm for analysis of AQI at DG Khan region.
3. To study and analysis of NO2 and O3 main factor of air quality index.

This work is organized as follows: Section 2 contains the literature review. Section 3 presents the methodology of the study. Section 4 discusses the results. In last, the conclusion and future work is described in the section.

2. Literature Review

Our contemporary environmental challenges include resource depletion, air pollution, global warming, hazardous waste, and many others. Millions of people die each year from illnesses brought on by exposure to outdoor air pollution (Alzubaidi et al., 2021) [5]. The risk of adverse health effects is growing over time, and particle overexposure is now a significant public health concern in China. The economy and energy use in Beijing, the capital of China, have both grown significantly in recent years. As a result of air pollution, Beijing's foggy weather has become more severe due to this rapid growth (Bak et al., 2013) [6]. The weather and air pollution frequently coincide and last for a long time, but over the past two years, things appear to have improved. A crucial indicator for determining and evaluating the air quality in a specific location is the Air Quality Index (AQI). According to the Chinese Standard GB3095-2012, the six primary pollutants are used to calculate the air quality index (AQI) (Liang et al., 2020) [7]. Inhalable particles (PM10), ozone (O3), sulphur dioxides (SO2), nitrogen dioxides (NO2), and carbon monoxide are some of these(Maity et al., 2020) [8]. This index has six levels and a scale from 0 to 500 that measures the general quality of the air (Shaban et al., 2016) [9]. These levels, which are shown numerically, demonstrate the effect on human health and provide a helpful guide for persons engaging in outdoor activities. Data from the Environmental Protection Agency are used to generate the air quality index (AQI) (EPA). High numbers indicate bad air quality, while low numbers indicate excellent air quality, which has an impact on people's capacity to engage in outdoor leisure activities.

To properly control and address air pollution issues, time and effort will be required. Air quality predictions based on weather may help prevent harm from air pollution. Air quality forecasting must take place on time if government organisations and the general public are to be able to preventative measures and avoid major pollution incidents. For instance, a number of facilities in Beijing, including coal-fired power plants and coking mills, have been forced to temporarily shut down due to the anticipated air quality (Obraczka & Rahm, 2022) [10]. Artificial intelligence (AI) or machine learning is the study of the statistical models and algorithms that computer systems use to make judgments or predictions without receiving explicit instructions (Pianosi et al., 2016) [11]. Machine learning has become very popular because it can make accurate predictions quickly when dealing with large amounts of data. (Brunelle & Chandel, 2002)[12]. Big data refers to enormous datasets, some of which may contain significant amounts of unstructured data that need more intensive real-time analysis to fully understand their hidden values. With varying degrees of success, some researchers have predicted short- and long-term air quality using machine learning algorithms. (Pianosi et al., 2016) [13]. DG Khan's hourly concentration of suspended specific matter was predicted using a multilayer neural network, with a focus on the problem of low prediction accuracy. (Caladcad et al., 2020) [14].

Since it describes the air quality in a specific area, the Air Quality Index (AQI) could be seen as a channel of communication between weather agencies and the general public. It can be thought of as an all-encompassing measurement that represents the local air quality (Cheng et al., 2007) [15]. This symbol represents the health risks associated with air and particle pollution. Explain the air quality in a clear and understandable manner. It is an understandable strategy (Conticini et al., 2020) [16]. With the help of these indicators, the general public may keep track of local, regional, and national air quality without needing to understand the details of the underlying data (Dat et al., 2020) [17]. People should be aware of air pollution so they may take the necessary precautions to shield themselves from its harmful effects. The public should be made more aware of the effects of current air pollution exposure in order to modify attitudes and behaviours. Government policies are also in place. Despite the fact that the AQI alone only depicts certain aspects of air quality, colour schemes, graphs, names of air quality categories (such as good, medium, or bad), and other messages are used in conjunction with the AQI (Douglass et al., 2014) [18]. Describe the expected impact of the index at various levels and the precautions people can take to reduce their risk of being impacted (Röver et al., 2021) [19]. When this is done, the results almost always show which specific areas are most likely to become more dangerous (Gao & Zhou, 2013) [20]. Therefore, it is illogical to think that the same air quality index applies globally (Wang & Zuo, 2022) [21]. The "Pollutant Standard Index (PSI)" was developed by the United States Environmental Protection Agency (USEPA). It rates air quality on a scale of "0 to 500," with 100 representing compliance with the "National Ambient Air Quality Standard (Sako et al., 2022)[22]. Calculate the pollutant specific indices (PSI) (Garcia et al., 2016) [23], Although they are all based on European Directive 2008/50 / EC air quality limitations, European member states use the same methodology, with slight variations between different member states and within different boundaries (Stančić et al., 2022) [24]. The Common Air Quality Index (CAQI), created by the CITEAIR project in 2005 with funding from the INTERREG project of the European Union, was intended to allow for real-time comparisons of air quality in European cities. The Air Quality and Health Index (AQHI) was developed in Canada with a public health perspective to analyse local air quality. (Ghorani-Azam et al., 2016) [25].

Every day, the economy of water and air is growing. In a few countries, it has gotten worse. Cities from all around the world are making improvements and cutting back on water pollution in various regions of the world. If the air is safe to breathe, the Air Quality Index counts the amount of pollutants in the atmosphere (Gregório et al., 2022) [26]. As a result of global air inhalation, they pass away. Pakistan reported 2,000 fatalities due to fungal infections brought on by unsanitary respiratory illnesses in the previous year as a result of the poor air quality there (Haidar et al., 2012) [27]. According to the index, dangerous levels of air pollution are present in about two-thirds of Pakistan's major cities, including Islamabad, Karachi, and Lahore, where all sectors maintain very high, high, and low air quality levels(Hájek & Olej, 2015) [28].

The mean metal concentrations in the atmosphere of DG Khan are much higher due to human activity than those of the region and large European cities (Xiao et al., 2015) [29]. Industrial metals like lead and cadmium are linked because they both come from the same source, whereas industrial metals like iron, zinc, manganese, and potassium showed strong connections (Iskandaryan et al., 2020) [30]. The principal component and cluster analyses revealed that mineral dust, combustion activities, industrial operations, and vehicle emissions were the main sources of airborne particle pollution in the atmosphere. Comparative research shows that there are significant amounts of trace metals in the air. (Karppa et al., 2020) [31]. The underlying statistical data demonstrated that the components changed widely depending on the season under consideration. (Ko et al., 2022) [32]. Major sources of pollution in DG Automobile emissions, wind-blown soil dust, excavation activities, biomass burning, industrial and fugitive emissions have all been identified as sources of Khan's atmospheric aerosols. When compared to other areas, the concentrations of airborne trace elements in this area are typically very high, which suggests that the locals may be at risk for health issues as a result of these elevated concentrations of trace elements.

3. Materials and Methods

This section contains the methodology of the study.

3.1. Proposed Methodology

In (Figure 1) block diagram of proposed methodlogy is expain. ANN is a mathematical model inspired by the natural process. It has several nodes that are connected to each other using different parameters and layers. The connection weights in such scenarios are stored into inter node processing network. The training processes of the nodes have processed using learning ability of nodes and associated weights. The ultimate objective of an ANN is to solve complex computing problems. ANNs are considered to as a black box consisting information about nodes, sets of inputs and outputs. The training data includes examples of inputs and outputs; the outputs are generally generated from the group of nodes and associated nodes. This model has been used instead of other approaches because of their propensity for prediction, robustness in the face of noisy data, and ability to analyse non-linear systems. Fuzzy logic systems are designed to operate on non-probabilistic functions and solve problems with fewer computing resources. Defuzzification strategies are used to preserve the interface engine, input, and output functions. Five sensors are employed as the core components of the NF-based air pollutant measure to measure the level of air pollution in the surrounding area.

In figure 1 block diagram explain five different sensor and it workingS:

1. β ray absorption method.
2. Non-dispersive UV absorption sensor.
3. The ultraviolet fluorescence method.
4. Non-dispersive infrared (NDIR) functions.
5. Gas Phase Chemiluminescence

These sensors are used to monitor pollutants like SO2, NO2, CO, PM2.5, O3 of the fuzzy controls system. This sensing system is attained by deploying hybrid NF learning mechanism. Each of the diffuser is connected via AQI indicator.

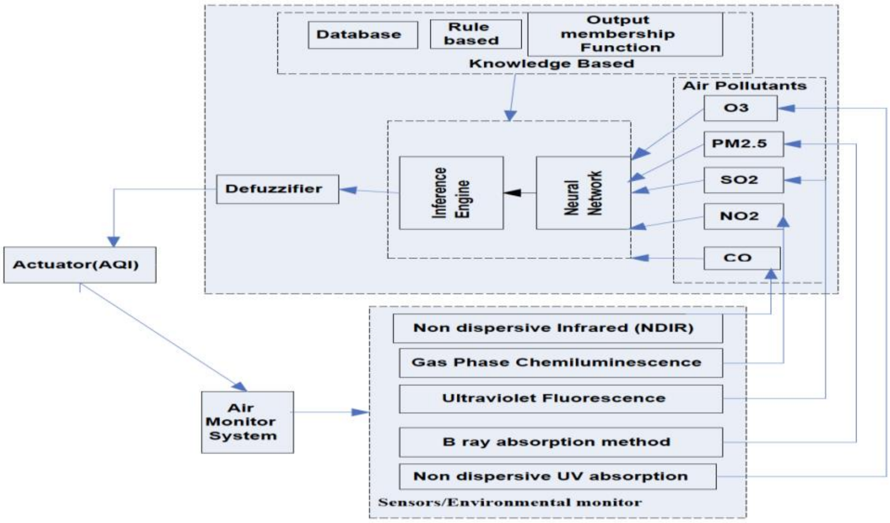

Figure 1. Block Diagram of Proposed Methodology

PAK-EPA has developed standards for National air quality and protection of human health. In this research, 24-hour analysis of the air pollutants were measured. The data was provided by EPD like 24-hour data, weekly data, and monthly data. The collected data was obtained when there were high-levels of the air pollution in the environment. From (table 1) analysis of Variance (ANNOVA) test was performed for training and test of the data set. The tested hypothesis reveals that each of the parameter discussed in this research has equal and significant value.

Table 1. ANOVA Training and Test Data

Groups	Sum of Squares	DF	Square Mean	F	Significance
Among Groups	2475285.513	4	618821.378	108.548	.000
Within Groups	5672380.724	980	6800.760	-	-
Total	9278563.421	996	-	-	-

From (table 2) analysis of Variance (ANNOVA) test was performed for training and test of the different data set.

Table 2. ANOVA Test for Training Data

Groups	Sum of Squares	DF	Square Mean	F	Significance
Among Groups	2917260.375	4	729315.094	132.794	.000
Within Groups	2718576.964	495	5492.075	-	-
Total	5635837.340	499	-	-	-

From (table 3) air pollution in this research is considered as inputs to the model and their concentrations over 24 hours. The selected pollutants for this research are: SO2 a reactive and colorless gas that is emitted as a result of fuel burning of Sulphur containing fuels. The major basis for the emission of this gas includes powerplants, refineries, and industrial boilers. PM2.5 is the smallest particle and commonly its size is 2.5 millimeters. Its sources include forest fires, use of vehicles, power plants, industrial, and combustion process. The concentration of ground level ozone is usually measured on hourly basis owing to change in vehicles' fuel consumption, cars, and chemical plants. O3 pollution is highest in the summer months. The ultimate result of the high-level of surface ozone is several respiratory diseases. CO is colorless and odorless gas and it is considered as one of the most fatal gases present in the environment. These pollutants have 24 hours impact on the air quality index.

Table 3. Build Breakpoints of Air Pollutants according to PAK-EPA Standards

O3 1hr	PM2.5 24hrs hr	CO 8hrs	SO2 24hrs	NO2 24hrs	AQI Category
1-55	1-19.0	1-3.5 Favorable	1-65	1-40	Favorable
1-135	0-40	1-6 Moderate	0-125	0-85	Moderate
Moderate	16.5-53.5	3.5-8.5 Unhealthy	0-130	45-135	Unhealthful
55-185	45-75	6-12 Hazardous	125-240	80-120	Severe

Table 4 explain break points for DG kjan for air pollution that is considered as inputs to the model and their concentrations over 24 hours.

Table 4. Break Points for DG Khan.

O3 1hrµg/m3	PM2.5 24 hr sµg/m3	CO 8hrs mg/m3	SO2 24 hr sµg/m3	NO2 24 hr sµg/m3	AQI Category
1-55	1-18.5	1-3.5 Favorable	1-65	1-40	Good
1-135	0-40	1-6 Moderate	0-120	0-80	Moderate
55-185	16.5-53.5	3.5-8.5 Unhealthy	60-180	40-120	Unhealthful

120-280	45-75	6-12 Hazardous	120-240	80-120	Severe

An integration system called ANFIS uses neural networks to enhance the fuzzy inference system. To create the specified input-output pairings, ANFIS builds a set of fuzzy if-then rules with the required membership functions. (Figure 2) based on the analysis so far, following ANFSI model can be established:

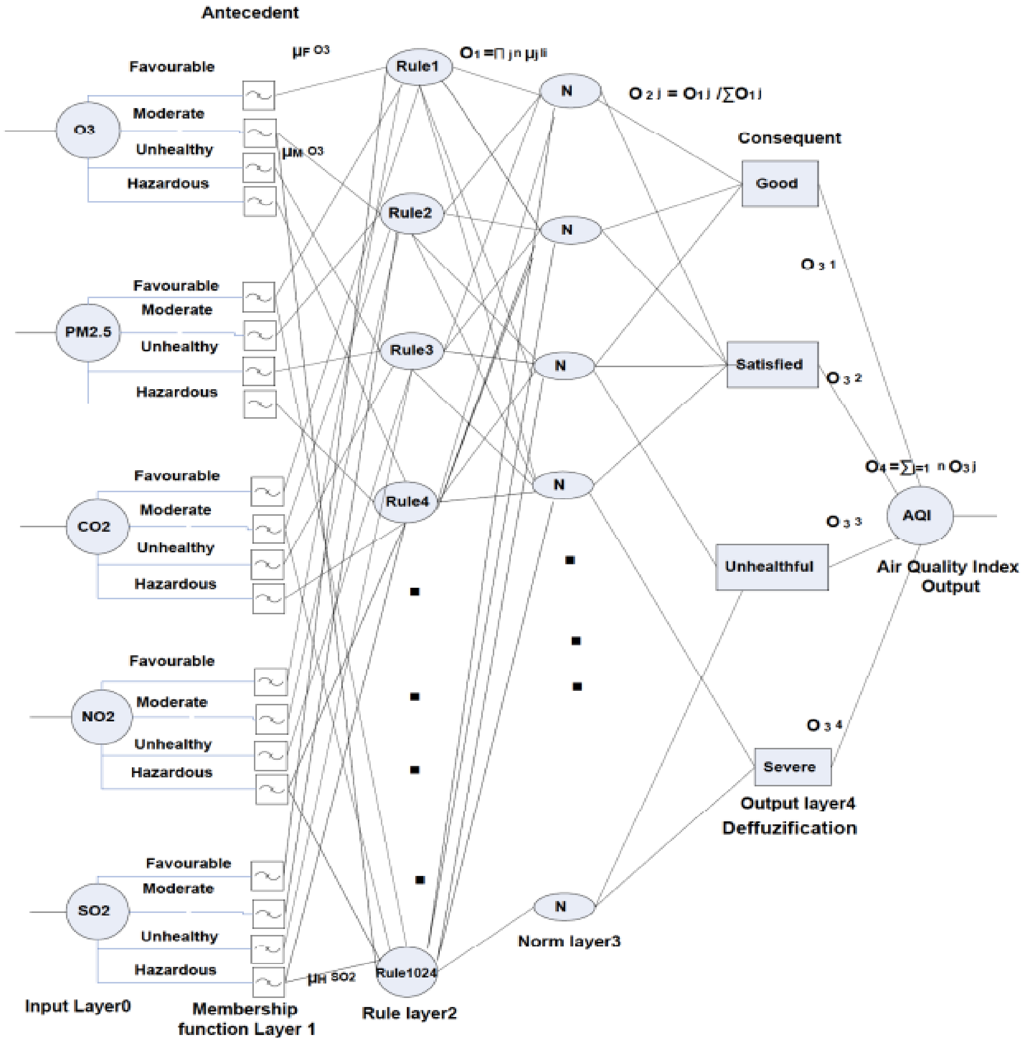

Figure 2. ASFI Model

In input layer 1, the crisp of input signal was transmitted to output signal. The governing equation is:

$$ab_{(i,j)}=O_{(i,j)} \tag{1}$$

In the 2nd hidden layer, also called fuzzy layer, the neurons represent themselves as a part of fuzzy network. The membership function for this operation is given by:

$$O_{(2,I)}=\mu A_i\left(ab_{(i,2)}\right) \tag{2}$$

In the equation above, ab is the neuron i, input function, and Ai is a label for food that includes words like "favourable," "moderate," "unhealthy," and "dangerous." The triangle membership of the function is established on the, which stands for activation. To calculate the three corners, these triangle parameters have three parameters: (a, b, c), and x. Different circumstances, such as favourable, moderate, unhealthy, and hazardous, have been represented by the four triangular member functions. In levels 1 and 2, respectively, of the aforementioned figure, these are connected as lines.

The second hidden layer in layer 3 is the ANFIS architecture, which is also referred to as the fuzzy rule layer. Each neuron in this configuration uses the first Sugeno fuzzy layer. The aggregation of neurons can be represented by the following equation:

O_(3,i)=w_i= 〖uA〗_i× 〖μB〗_i×(〖SO〗_2)×μC_i (〖PM〗_2.5)×μD_i (〖NO_2)× 〖μD〗_i×μ_i (CO) (3)

If the gases are in favorable conditions that it can be assumed that Air Quality Index is fine. If one these gases are not in favorable condition that the output for AQI will be moderate. The proposed is designed for the prediction of AQI and has shown significant performance using NF performance index. ANFIS MATLAB functions are used to compared to conventional linear interpolation model. There were almost 200 samples analyzed during this analysis, from which, almost 100 samples were trained for the testing and training of the data. The 0.5 was kept as a tolerance value. The air mass trajectory mechanism has been widely used to propagate the direction oi wind. The prediction of PM2.5 in higher concentration of air is of vital importance to forecast models and formation of both forward backward trajectory.

4. Results and Discussions

The particles found in the atmosphere have different types and composition. According to estimates a particle having 1 μm must have some sort of internal properties. The typical suspended particles have some composition of the inorganic materials like SO2-4, nitrates, and compounds including levels of sodium, chloride, and ammonia. There are some regulations that can limit the presence of suspended particles to PM2 to PM10 indication the values should be between 2.5 μm to 10 μm. In the cities like DG khan the values like μg-3. PM loading in the atmosphere adds to the loss of visibility associated with poor air quality, and has both direct and indirect effects.

Climate change is caused by the transmission of solar energy through the atmosphere. These impacts (in nature, primarily but not solely cooling) one of the most significant and poorly measured concepts in climate science models and might have a significant impact on the calculated warming's magnitude. Particles of aerosol also play a vital part in the chemistry of the atmosphere the provision of a reaction location for heterogeneous reactions would not have happened otherwise. The transfer of dust is crucial component of the crustal biogeochemical cycles' minerals, particularly iron, are an important source of income for them contribution to marine ecosystems.

Predictive modelling, often known as regression, is a method for fitting data to a specified range of values. Python's regression methods are described in detail here:

- Random Forest Regressor
- Ada Boost Regressor
- Bagging Regressor
- Linear Regression etc.

Table 5. Excel file with a sample of the dataset

Dera Ghazi Khan District	Year	Month	Date	O3	NO2
Rajanpur	2021	September	18	1439	35
Muzaffargarh	2021	September	18	1955	57
Layyah	2021	September	18	4083	714

DG Khan	2021	September	18	3067	24

From (table 5) air quality index (AQI) captures the daily pollutant parameters in accordance with the DG Khan City Standard (2021). The Excel dataset has been enhanced by the addition of a new attribute column that has a binary categorization (The Area is Polluted or Not Polluted). There was an additional column added to indicate the name of the most effective pollutant and, therefore, the categorization values. In addition, one of the two contaminants or sources of contamination should be labelled. The characteristics are independent factors, and they all have an impact on the dependent variables (class label), which are also independent variables (pollution column).

4.1 Analysis

The predictive modelling of atmosphere and pollutants is most of the times done by the use of ANN. These models are effective in discussing and prediction of ozone layer's depletion. As a result, most of the research works focus on the development of ANN and SVM mechanisms to address the concentration of air pollutants in the environment. In the current research, ANN was used to compare and analyze different pollutants in the air. However, in the current research a special focus was also given to the use of statistical tools like ANNOVA analysis. However, findings suggest that Neural Networks work best when dealing with non-linear datasets. For most of the historical analysis the research works have focused on their studies on pollution monitoring and analysis, use of non-linear methodologies, support vector machines, and ANN. It should be noted that most of these methods require necessary training before deploying in the real-time environment. The different concentrations of the air pollutants in the environment and learning mechanism are based on the established parameters and use of conventional software packages. There are few studies that have restrictions imposed by small number of alternative software products and methods. Therefore, it can be observed that there is no reliable study available to accurately predict and model the surface level O3.

The use of time-series data is considered as one of the most reliable forms of statistical analysis. It has been used in wide range of applications during the recent decade. The analysts perform their predictive analysis based on the historical data available to them. This prediction helps them to forecast future behavior of the data and associated parameters. The effective use of time-series data can facilitate the in educating the general public about potential hazards, give early warnings, and take necessary actions to limit the effect of hazards on the population.

The time series data consists of a collection of data gathered at regular intervals (monthly, daily, hourly) basis. This data can be discrete or continuous in nature. The data in such cases is time dependent and sequence of representation is of extreme importance for this type of analysis. The current research used ANN empowered by time series data to predict the levels of pollutants like O3, NO2, SO2, and others. Three types of pollutants are most commonly known as agricultural, industrial, and residential. Inhaling these air pollutants can prove hazardous for the individuals. There are several individuals that suffer from fatal consequences by inhaling such pollutants.

Shortness of breath, eye etching, irritation of nose, eyes, and throat are common symptoms of bad air quality. It is also observed that poor air quality can also result in cardiovascular disorders. If the patient is already suffering from respiratory illness; it can become severe. Despite worsening physical effects, the inhaling of pollutants can also have negative consequences for the psychological development. It is observed that high-level of exposure to harmful gases can cause psychological discomfort.

An important thing to notice in this scenario is the fact that excessive physical activity has limited effect on the inhaling of harmful gases. However, it is observed that greater physical activity can cause severe problems in case of poor air quality. An outdoor physical activity put the users under more physical and psychological strain. Another danger that stems from the inhaling of polluted air is the negative consequences suffered by the children. Since children are smaller it can be estimated that they inhale less air compared to adults. Children are therefore, more vulnerable to the air pollution. Furthermore, children spend more time outside compared to adults and therefore can become a direct victim of the poor air quality. The fragile ecosystem of the plane is also vulnerable to the air pollutants. Airborne pollutants can have negative consequences for normal plant and crop growth.

5. Conclusion and Future Work

It is difficult to predict air quality because of the dynamic nature, volatility, and high unpredictability of pollutants and particles in space and time that characterize the environment. Simultaneously, being able to accurately estimate, forecast, and monitor air quality is becoming more essential, particularly in metropolitan areas, as a result of the known negative effects of air pollution on both human health and the environment. As a result of this research, it was possible to predict pollutant and particle levels as well as accurately identify the air quality index (AQI) using support vector regression (SVR). It was discovered that the investigated approach generated a reasonable model of hourly atmospheric pollution, which allowed us to achieve usually excellent precision when modelling pollutants such as O3, CO, and SO2, as well as the hourly air quality index (AQI) for the state of Californi. As part of our future work, we aim to enhance and explore the use of SVR to predict air quality by focusing on the following areas: Selection of data sets and variables In the case of NO2 and PM2.5 a big dataset with a greater number of factors and observations may be used to develop more accurate prediction model. Fifth, as the choice of the kernel function and the penalty parameter have a significant impact on the SVR model's performance, it would be interesting to look at different approaches to hyperparameter optimization, including genetic algorithms or particle swarm optimization. In addition to random search. Most importantly, we aim to compare the results produced by SVR to those obtained by several other machine learning algorithms of varying type, such as artificial neural networks, Bayesian network models, decision trees, random forests, and genetic programming.

References:

1. Adhikari, R., & Agrawal, R. K. (2013). An Introductory Study on Time Series Modeling and Forecasting. ArXiv, abs/1302.6613.
2. Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. IEEE Transactions on Knowledge and Data Engineering, 17(6), 734-749. https://doi.org/10.1109/TKDE.2005.99
3. Afroz, R., Hassan, M. N., & Ibrahim, N. A. (2003). Review of air pollution and health impacts in Malaysia. Environ Res, 92(2), 71-77. https://doi.org/10.1016/s0013-9351(02)00059-2
4. Altaheri, H., Alsulaiman, M., & Muhammad, G. (2019). Date Fruit Classification for Robotic Harvesting in a Natural Environment Using Deep Learning. IEEE Access, 7, 117115-117133. https://doi.org/10.1109/ACCESS.2019.2936536
5. Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., & Farhan, L. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. Journal of Big Data, 8(1), 53. https://doi.org/10.1186/s40537-021-00444-8
6. Bak, J., Kim, J. H., Liu, X., Chance, K., & Kim, J. (2013). Evaluation of ozone profile and tropospheric ozone retrievals from GEMS and OMI spectra. Atmos. Meas. Tech., 6(2), 239-249. https://doi.org/10.5194/amt-6-239-2013
7. Brunelle, J. K., & Chandel, N. S. (2002). Oxygen deprivation induced cell death: An update. Apoptosis, 7(6), 475-482. https://doi.org/10.1023/A:1020668923852
8. Caladcad, J. A. A., Cabahug, S., Catamco, M. R., Villaceran, P. E., Cosgafa, L., Cabizares, K. N., Hermosilla, M., & Piedad, E. J. (2020). Determining Philippine coconut maturity level using machine learning algorithms based on acoustic signal. Comput. Electron. Agric., 172, 105327.
9. Cheng, W. L., Chen, Y. S., Zhang, J., Lyons, T. J., Pai, J. L., & Chang, S. H. (2007). Comparison of the Revised Air Quality Index with the PSI and AQI indices. Sci Total Environ, 382(2-3), 191-198. https://doi.org/10.1016/j.scitotenv.2007.04.036
10. Conticini, E., Frediani, B., & Caro, D. (2020). Can atmospheric pollution be considered a co-factor in extremely high level of SARS-CoV-2 lethality in Northern Italy? Environ Pollut, 261, 114465. https://doi.org/10.1016/j.envpol.2020.114465
11. Dat, N. Q., Thi, N. A. N., Solanki, V. K., & Le An, N. (2020). Prediction of Water Level Using Time Series, Wavelet and Neural Network Approaches. International Journal of Information Retrieval Research (IJIRR), 10(3), 1-19. https://doi.org/10.4018/IJIRR.2020070101
12. Douglass, A. R., Newman, P. A., & Solomon, S. (2014). The Antarctic ozone hole: An update. Physics Today, 67(7), 42-48. https://doi.org/10.1063/PT.3.2449
13. Gao, W., & Zhou, Z.-H. (2013). On the consistency of multi-label learning. Artificial Intelligence, 199-200, 22-44. https://doi.org/https://doi.org/10.1016/j.artint.2013.03.001
14. Garcia, J. M., Teodoro, F., Cerdeira, R., Coelho, L. M., Kumar, P., & Carvalho, M. G. (2016). Developing a methodology to predict PM10 concentrations in urban areas using generalized linear models. Environ Technol, 37(18), 2316-2325. https://doi.org/10.1080/09593330.2016.1149228
15. Ghorani-Azam, A., Riahi-Zanjani, B., & Balali-Mood, M. (2016). Effects of air pollution on human health and practical measures for prevention in Iran. J Res Med Sci, 21, 65. https://doi.org/10.4103/1735-1995.189646
16. Gregório, J., Gouveia-Caridade, C., & Caridade, P. J. S. B. (2022). Modeling PM2.5 and PM10 Using a Robust Simplified Linear Regression Machine Learning Algorithm. Atmosphere, 13(8), 1334. https://www.mdpi.com/2073-4433/13/8/1334
17. Haidar, A., Dong, W. H., & Mavridis, N. (2012). Image-based date fruit classification. 2012 IV International Congress on Ultra Modern Telecommunications and Control Systems, 357-363.
18. Hájek, P., & Olej, V. (2015). Predicting Common Air Quality Index-The Case of Czech Microregions. Aerosol and Air Quality Research, 15, 544-555.
19. Iskandaryan, D., Ramos, F., & Trilles, S. (2020). Air Quality Prediction in Smart Cities Using Machine Learning Technologies Based on Sensor Data: A Review. Applied Sciences, 10(7), 2401. https://www.mdpi.com/2076-3417/10/7/2401

20. Karppa, M., Kaski, P., Kohonen, J., & P, Ó. C. (2020). Explicit Correlation Amplifiers for Finding Outlier Correlations in Deterministic Subquadratic Time. Algorithmica, 82(11), 3306-3337. https://doi.org/10.1007/s00453-020-00727-1
21. Ko, H., Lee, S., Park, Y., & Choi, A. (2022). A Survey of Recommendation Systems: Recommendation Models, Techniques, and Application Fields. Electronics, 11(1), 141. https://www.mdpi.com/2079-9292/11/1/141
22. Liang, Y.-C., Maimury, Y., Chen, A. H.-L., & Juarez, J. R. C. (2020). Machine Learning-Based Prediction of Air Quality. Applied Sciences, 10(24), 9151. https://www.mdpi.com/2076-3417/10/24/9151
23. Maity, B., Polapragada, Y., Ghosh, A., Bhattacharjee, S., & Nandi, S. (2020). Identifying Outdoor Context by Correlating Air and Noise Pollution Sensor Log. 2020 International Conference on COMmunication Systems & NETworkS (COMSNETS), 891-893.
24. Obraczka, D., & Rahm, E. (2022). Fast Hubness-Reduced Nearest Neighbor Search for Entity Alignment in Knowledge Graphs. SN Computer Science, 3(6), 501. https://doi.org/10.1007/s42979-022-01417-1
25. Pianosi, F., Beven, K., Freer, J., Hall, J. W., Rougier, J., Stephenson, D. B., & Wagener, T. (2016). Sensitivity analysis of environmental models: A systematic review with practical workflow. Environmental Modelling & Software, 79, 214-232. https://doi.org/https://doi.org/10.1016/j.envsoft.2016.02.008
26. Röver, C., Bender, R., Dias, S., Schmid, C. H., Schmidli, H., Sturtz, S., Weber, S., & Friede, T. (2021). On weakly informative prior distributions for the heterogeneity parameter in Bayesian random-effects meta-analysis. Research Synthesis Methods, 12(4), 448-474. https://doi.org/https://doi.org/10.1002/jrsm.1475
27. Sako, K., Mpinda, B. N., & Rodrigues, P. C. (2022). Neural Networks for Financial Time Series Forecasting. Entropy, 24(5), 657. https://www.mdpi.com/1099-4300/24/5/657
28. Shaban, K. B., Kadri, A., & Rezk, E. (2016). Urban Air Pollution Monitoring System With Forecasting Models. IEEE Sensors Journal, 16(8), 2598-2606. https://doi.org/10.1109/JSEN.2016.2514378
29. Stančić, A., Vyroubal, V., & Slijepčević, V. (2022). Classification Efficiency of Pre-Trained Deep CNN Models on Camera Trap Images. Journal of Imaging, 8(2), 20. https://www.mdpi.com/2313-433X/8/2/20
30. Wang, L., & Zuo, K. (2022). Medical Data Classification Assisted by Machine Learning Strategy. Computational and Mathematical Methods in Medicine, 2022, 9699612. https://doi.org/10.1155/2022/9699612
31. Xiao, H., Biggio, B., Nelson, B., Xiao, H., Eckert, C., & Roli, F. (2015). Support vector machines under adversarial label contamination. ArXiv, abs/2206.00352.
32. Yu, R., Yang, Y., Yang, L., Han, G., & Move, O. A. (2016). RAQ-A Random Forest Approach for Predicting Air Quality in Urban Sensing Systems. Sensors (Basel), 16(1). https://doi.org/10.3390/s16010086