Nanotechnology Perceptions
ISSN 1660-6795
www.nano-ntp.com

A Pilot Study for the Detection of Air Pollution of Puducherry, India by Machine Learning Algorithms

V. Priya[1], S. Neduncheliyan[2*]

[1]*Research Scholar, Bharath Institute of Higher Education and Research, Chennai, Tamilnadu. priya.v@grt.edu.in*
[2]*Professor/Dean, School of Computing, Bharath Institute of Higher Education and Research, Chennai, Tamilnadu, dean.cse@bharathuniv.ac.in*

Air pollution has been the most pressing issue in human evolution in the last century, since high levels of hazardous chemicals and particles in the atmosphere cause health problems and have an impact on the planet's ecosystems. As technology advances, Engineers and Researchers of environment keenly exploring for new ways to prevent and regulate air pollution, resulting in innovative solutions. The Internet of Things (IoT) has inspired new approaches to pollution management, with instruments or sensors which have the potential to observe, forecast and offers regulatory direction were becoming more accessible and less expensive during the last decade. Air quality monitoring is a required task in many industrial and urban regions across the world. In places with substantial air pollution concerns, Air Quality Indexes (AQIs) are produced for this purpose. The AQIs are operational units in charge of administering monitoring networks, processing data, and ultimately giving on-line assessments of air pollution and its short- and long-term trends. These centres capture massive amounts of data throughout time. One of the main concerns of scientists working on AQIs is the availability of appropriate modelling tools for interpreting and validating the data produced. In this case, machine learning approaches appear to be appropriate. The goal of the study described here is to see if machine learning techniques can be used to anticipate air quality in real-world situations. The use of machine learning algorithms under operational air quality monitoring situations for estimating the daily peak concentration of a significant photochemical pollutant - NO2- from point data of a local monitoring station is the subject of this article. Our analysis relied on data from the AQI of Puducherry, India. The author introduces and discusses a Case-Based Reasoning system designed for air quality monitoring. Its performance is compared to that of a decision tree and a back propagating neural network technique (CART). The use of these algorithms in a normal AQI on a day-to-day basis is also explained.

Keywords: Air pollution, Internet of Things (IoT), Air Quality Index, major photochemical pollutant - NO2, neural network algorithm, Machine Learning.

1. Introduction

Pollution is, without a doubt, the most serious issue of the previous decade, one that is regularly debated in the media, in government meetings, and in environmental activities. Air pollution has been hard to manage and control for decades, yet it can be combated with the assistance of technological developments. Using the IoT and ML technology, pollution might be monitored and regulated, as well as forecast future increases in air pollution in urban areas [1]. The construction of an index to predict and assess the quality of air using sensor technologies is required for monitoring air pollution levels. The Air Quality Index (AQI) is calculated, so getting an perception of the sources of pollutants of a particular area based on maximum presence of NO2 symbolizes the usage of fossil fuel by combustion of vehicles in traffic zones of the city, which could be due to heavy traffic, for example [2].

The AQI, also known as the Air Pollution Index (API) or Pollutant Standard Index (PSI), produces a picture of air quality in a certain range for each air pollutant. Various sensors from a variety of categories might be utilized to calculate the precise value of the AQI and to identify the pollutants suspended in the air were causing the disaster, such as electrochemical sensors were based on the chemo-reaction of the air and electrode within an aqueous medium in the sensor, even optical sensors or optical particle counters or photo-ionization detectors [4].

Photochemical air pollution is a serious environmental issue in many large cities. This is a natural phenomena caused by a chain of chemical reactions driven by sunlight. The photochemical process is made up of nitrogen oxide, nitrogen dioxide (NO_2), ozone (O_3), and energy derived from sun light. Several studies have shown that nitrogen dioxide and ozone pose substantial health risks to the general populace.

We developed a system that uses pollution data from government locations and static sensors to produce a Machine Learning model. The learned module used to estimate the quality of air in Pondicherry at any given hour or day. It shows that constructing a Predictive Model for forecasting NO_2 levels is feasible.

Until date, air pollution models have mostly relied on dispersion models, which provide a rough representation of the complicated physicochemical processes involved. While the intricacy and complexity of these models has grown over time, their usage in real-time air pollution monitoring appears to be unsuitable in terms of performance, input data needs, and adherence to the problem's time limitations. Instead, human specialists' expertise has been employed mostly in Air Quality Operational Centers for real-time judgments, while mathematical models have been used primarily for off-line investigations of the phenomena involved". It seems that Machine Learning (ML) approaches, as an active do- main of artificial intelligence, can play a supportive role in such scenario. These techniques take advantage of the vast amounts of available data and their capability to process uncertain and incomplete information, providing fast data to give a real-time forecast.

2. System Design and Implementation

As a result, automation of the process has become necessary. This is expected to happen as part of a planned integrated system that gathers all essential data, estimates the risk of various pollutants causing an incident, forecasts short- and long-term trends, and proposes remedies as needed. The structure of this in-formation centre will be discussed next. "Multiple inputs are included in the operating center's design, including many from the air quality monitoring network of stations, others from the national meteorological service's

(EMY) weather bulletins, and yet others from external sources, such as traffic control data. Figure 1 depicts the task of air pollution forecasting.

If the quick classification model forecast for an air pollution event is positive, the accurate prediction method, which is based on complicated mathematical models, is initiated, according to the recommended air quality prediction strategy for Athens AQI.

Using the obtained air quality forecast and air emissions inventories, scenarios of obligatory countermeasures aimed at reducing emissions might be developed. The model is then informed that the next forecast of air pollutant levels may begin, making it feasible to estimate emissions. Furthermore, pollutant emission predictions can be utilised to assess the efficacy and societal effect of essential rectification actions. There may be a selection of additional scenarios based on the outcome of the review.

Source of Data

The following two sources of information were used to compile the data:

1. RTC Module, SD Card Module, Arduino Uno, Static Sensor Circuit & MQ7 Sensor

2. Open access Source Government Website www.data.gov.in

Sequencing of collected Data

Upcoming strategy will be divided into 2 sections, as follows:

(i) Preparation of Target: This stage entails transforming data into a format that MLM Machine Learning models can understand. This is accomplished by using one of the various target preparation methods available, such as Max- Min Normalization which converts data to fall between any given interval, Standardization which transforms data to have a mean of 0 and the same standard deviation and so on.

(ii) Preprocessing Data: This stage entails sequencing the primary data by converting one feature to another, eliminating null values extracting additional features from one feature, undesirable values and so on...

The parameters are estimated using the K-fold cross validation method parameter tuning. At random, the original data sample is split into k groups of equal size. The first group is utilised to test the model, while the following data sequenced from the k-1 groups will be used as the set of training data sets.

The user interacts with the programme using a basic command line interface, where the user enters the day, which ranges from A to G, with A being Sunday and G being Monday, and the hour of the day, which ranges from 00 to 24.

The initial stage in utilising machine learning algorithms is to prepare the data; the data was split into 60 percent data training & 40 percent test data for this purpose. Our data covers 210 admissions between April 2020 and March 2021, when Corona virus lockdown restrictions in Puducherry were at their greatest and open circulation was prohibited and 120 admissions between January 2021 and December 2022, when restrictions were lowered depending on the virulence. Heat in the form of Temperature in Centigrade, NO2 quantity, SO2 quantity, CO quantity, all were split into five columns. Following that, an appropriate method must be chosen; Artificial neural networks have been used by some researchers, while machine learning approaches have been used by others [10-17]. Random regression forest method is the most commonly utilised machine learning approach in earlier papers; it works by training numerous decision trees simultaneously. A decision tree separates data into categories and eliminates any distinctive factors that may be further classified [18-24]. Each decision tree is handled as a distinct class, with each tree being independently to the others. Because it is a fundamental component of the Random Forests approach, decision-tree learning is easy to utilise on its own [25].

This method uses a single decision tree to map the whole set of collected data, predict outcomes, and make observations. Another popular method is to employ Support Vector Machines, which can do both regression tasks and classification. SVMs divide data along a hyper plane, which serves as a classification boundary. SVMs that are closest to the hyper plane and are equidistant from it. Any change to these support vectors causes the categorization to change [24]. The ability to respond to data acquired by sensors is one of the aims of smart cities. However, it is important to note that sensors may cause inaccuracies and data failures; hence, in order to manage air quality, a model to forecast the values of interest must be developed [13]. One of the six atmospheric stations in Chennai, India data was collected. Each station is strategically located in the city's most congested areas, where a large number of people commute and live. Every hour, data on meteorological factors and chemical constituents present in the air were gathered. The stations may differ in terms of equipment, but they all provide vital information.

The data gathered and processed using the techniques described above may be utilised to construct weather and AQI forecast. IoT technology has become the important view in the perception by Algorithm developers, corporations & engineers since we live in a technological era when every gadget is connected to the internet, regardless of its size or function. By linking the system described in this study to traffic cameras and larger causes for pollutants, likely industries and unwanted burning of wood materials, volcanic eruptions, a comprehensive system capable of warning and combating air pollution may be created. For example, if a spike in NO2 particles is detected near an atmospheric station, one can deduce that these particles are the result of burning. Based on these facts, traffic congestion is likely to occur near the station. The device can produce a map of likely pollution sources by measuring wind speed and direction. It may then use the cameras to visually hunt for the cause, eventually alerting authorities to a potentially discovered fire.

The fast prediction model must comply with the following requirements:

1. It must have as inputs data from multiple sources, such as historical air quality measurements, meteorological forecasts, pollutant emissions data, episode levels definition, historical measurements of surface and upper air meteorological data,
2. It must have been successfully applied in the past to an urban area similar to that of the application area,
3. It must have accuracy prediction rate very high, so that it can be used in operational basis, it is expected that this value should be greater than 70%,
4. It must incorporate predictions for all major pollutants, like NO2, O3, CO, SO2 and smoke,
5. It must give predictions for the following three different time windows; 12 hour, 24 hour and 48 hour predictions,

6. It must provide a quick response, e.g. maximum prediction time not higher than 15 minutes,
7. It must have an easy to use interface and must present the results in an understandable way to non-computer experts.

The system may also use traffic cameras to verify that this is the origin of the pollutants. If so, the system can begin looking for a way to manipulate the traffic lights in order to increase circulation, eventually alerting other channels of congestion in traffic to choose an another route. Figure 2 depicts the fundamental diagram of such a system. Another example is when excessive NO2 levels are discovered. This might also signal the presence of a fire near the station. The particles' source may be anywhere within a reasonable radius. Wind sensors can assist in limiting the amount of space available by locating the source.

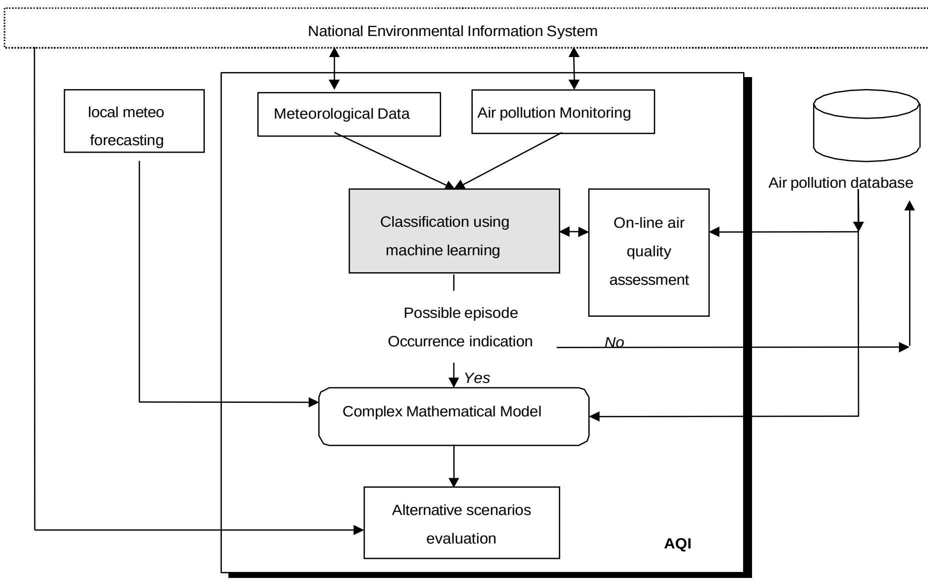

Fig. 1. Air quality prediction system architecture for AQI

3. The Implementation of The Machine Learning Methods

We tested 3 prediction algorithms coming from 3 active trends of machine learning research and practice. The first is a case-based reasoning (CBR) modified algorithm, the second is a multi-perceptron neural network, and the third one is a common decision tree approach. More details are given in the following about the CBR system than for the other two, since the CBR methodology is quite general, resulting to several different implementations of the CBR paradigm. On the other hand, the two inductive algorithms tested, i.e. the decision tree and the artificial neural net- work are well defined in theory.

Case-Based Reasoning – The NEMO System

The case-based reasoning method applies previous knowledge to answer new issues; the core concept underlying CBR is to employ a systematically documented repository of previous examples of the problem to be addressed. When the system is given with a fresh problem, it searches the database for the most comparable previous situations, which are then tailored to the current situation to produce the new answer [16].

Despite the fact that CBR systems and applications developed so far present several important functional differences, a general diagram of actions for the design of CBR systems has been widely accepted. Figure 2 shows the basic phases of this process.

The CBR implementation named NEMO was based on a previous CBR system called AIRQUAP developed for the same application domain.

The knowledge base has replaced all measures modification heuristics, similarity metrics. We couldn't have a test solution step because of the nature of the problem, thus the diagram loop had to be streamlined. NEMO is made up of three primary modules, as can be seen from a broad perspective of the figure. Two of them are in charge of retrieving and filtering comparable cases to the new case, while the third alters the solutions supplied by the remaining similar cases to generate the proposed solution for the new case [21]. In its current state, NEMO can forecast the maximum NO2 concentration in a measurement station region that has been previously established.

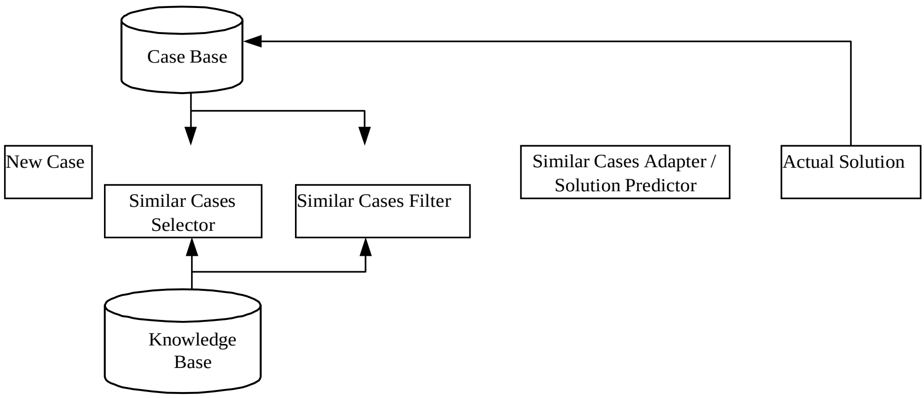

Fig. 2. Block diagram of Case Based Reasoning Module NEMO

4. Related Works

D.J. Briggs et al. [1] developed a geographic information system (GIS)-based approach for mapping traffic-related air pollution. NO2 was taken into account at Huddersfield, Prague, and Amsterdam. Their findings show that they can predict pollution levels properly.

V. Singh et al. [2] have presented a technique for estimating and interpolating daily ozone levels. This method uses a process known as cokriging. M Mead et al. [3] placed sensor nodes in a static network and a mobile network in the Cambridge (UK) region. They have supplied the findings for personal exposure quantification.

Individuals can assess their own exposure using participatory sensors, and organisations can summarise their members' exposure, according to P. Dutta et al. [4]. K Hu et al. [5] used a variety of software tools and hardware devices to collect high-density urban air pollution data.

Ke Hu et al. created a machine learning model that leverages data from fixed stations and mobile sensors to measure air quality in Sydney at any hour of the day. They employed ten-fold cross validation and seven regression models [10].

Kumar Saha et al. [12] used a cloud-based Air Pollution Monitoring System run by a Raspberry Pi. They computed the Air Quality Index using a MQ135 and Gas Detection Sensor Air Quality and five criteria pollutants:, Carbon Monoxide, Nitrogen Dioxide, Sulphur Dioxide, ground level ozone and particulate matter.

Pondicherry has a population density of 1,642 people per square kilometre. It also has some of the country's densest air. Maximum levels of spm and SO2 were roughly three times the acceptable limits in 1997, while NO2 levels were more than "two times the permissible limit." Pondicherry is one of the few cities - and the only one among the Down To Earth case studies - where all of the metrics' highest values were well beyond the permissible limit in 1997. This has been the pattern since 1988, when the monitoring began. Pondicherry's industrial expansion has been completely uncontrolled, and its industrial policy was officially proclaimed for the first time in 1997. By that time, a lot of harm had already been done. As of September 1998, there were 6,038 industrial units in Pondicherry, of which 136 are in the large- and medium-scale sectors. The number of vehicles registered in Pondicherry has almost doubled from 79,748 in 1991 to 152,719 in 1998. “In the same period, the number of vehicles for transport of goods, which mostly run on diesel, almost tripled - from 1,794 to 5,000. Not surprisingly, air pollution has also increased. What makes the situation worse are malpractices such as adulteration of fuel.

Training the other ML methods

In our ANN experiment in figure 3, we used a feed forward multi-perceptron network with 1 output node, 2 hidden layers of 6, and 10 input nodes. The entire step functions at the hidden layer nodes are Gaussian. The training technique was error back propagation (McCelland Rumelhart 1986), with the network performing well against the training set after 5-6 hours of effort. Many less successful attempts have been been out, involving the use of networks with various topologies.

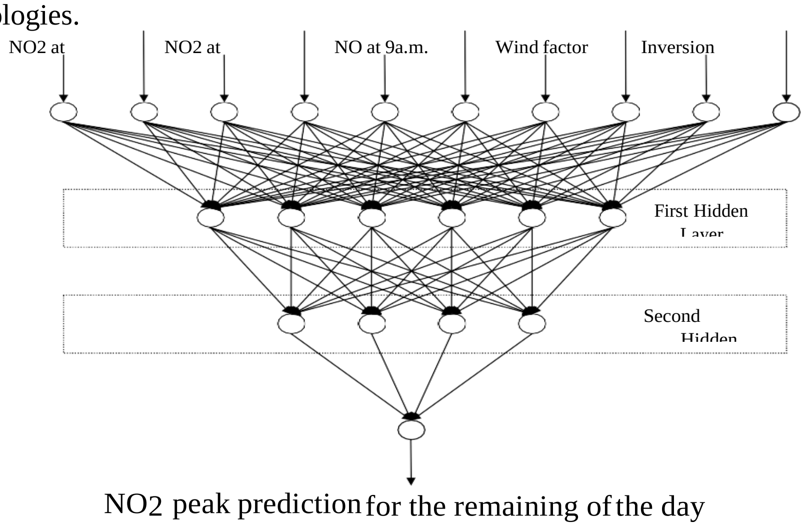

Fig. 3. The architecture of the ANN used for experimentation

In addition to the above methodologies, an inductive top down decision tree was utilized, namely the Oblique Classifier (OC1) built by S. Murthy (Murthy 1994), which has been shown to outperform typical decision tree algorithms such as ID3, C4.5 (Quinlan 1993), and its descendents. The fundamental point of OC1 is that the tree may divide at each node based on the algebraic sum of several qualities, rather than just one, as with typical C4.5 applications. The OC1 can produce axis parallel trees, CART-like trees (classification and regression trees) or trees made by oblique splits. It can handle missing values that are replaced by the average values of the corresponding attributes. Several missing values exist in our input files, due to occasional malfunctioning of the measurement appliances or due to communication problems that made impossible the data transfer from the scattered stations to the central monitoring station. All missing values in the data were intentionally not removed as they simulate typical operational conditions of an AQI and they are common in real monitoring applications.

Table 1. Levels of NO2 concentrations for the Puducherry AQI

level	In J P^3
1	0 - 100
2	100-250
3	250-600
4	over 600

The air quality monitoring authority for the area, AQI of Puducherry, has obviously been our primary source of information. This government agency operates a network of 12 air quality monitoring stations that are evenly spread throughout the area. The hourly mean readings of SO_2, NO, CO_2, NO_2, O_3, humidity, wind speed, temperature, and wind direction are all recorded by this network. Additional data was obtained from the national meteorological agency, known as EMY, which provides the area's meteorological prediction, as well as information on upper-atmosphere conditions, temperature inversion information, and so on.

From the EMY prediction bulletins, we retrieved all important information about temperature inversion below 150 m above the surface, temperature, future wind speed and direction. The National Observatory's recorded precipitation and sun radiation levels were added to this data. The data belongs to the expert observations of overall meteorological conditions in the area at run time.

We aggregated all of this information into a pollution data pool that spanned two years. There were still a number of issues with the data's quality and integrity. We had to deal with concerns of validity and integrity in order to generate a meaningful data set from this database, so we asked the experts and gathered some heuristic measures over each of the measurements to increase their quality. We also had to deal with the missing data. At this point, our objective was to create a final credible data set. We opted to include just the characteristics that were more important to NO_2 pollution levels in the final data set, and we turned some of them into less error-prone attributes, abstractions of the raw values, as explained below.

The NO and NO_2 mean hourly concentrations from 7 a.m. to 10 a.m., as well as a set of meteorological attributes such as the wind factor, which is a function of wind speed and direction, the temperature inversion factor, the precipitation factor, and the solar radiation factor, were all included in the test data set. These variables were calculated using the heuristic functions that were created, and they represent a measure of how well each element contributed to the progression of the NO_2 episode. For wind-related data, there are two sources: real field measurements taken by the AQI network and projections provided by EMY the national meteorological service. The wind factor derived from real observations is a combination of wind speed and direction.

In the evolution of air pollution events, the temperature inversion is beneficial. This occurs when a lower layer of air produces a temperature that is greater than the temperature of the upper atmospheric layers. This layer functions as a trap for pollutants, preventing them from entering the lower atmosphere. This phenomenon creates large concentrations of photochemical pollutants, especially when there is no breeze. A heuristic Rain factor is used to explain rain and associated occurrences. There are two possible origins.

Table 1 shows the four primary groups or tiers that the output attribute has been divided into. This restriction was imposed by the decision tree method since it is unable to give numerical continuous values as an output. Because both decision trees and neural networks approaches can't handle missing values, the attribute mean was used to fill them in. However, there are certain implementations that go around this issue. We chose the most important characteristics based on the physics of the photochemical process, as stated by environmental specialists. All NO_2 measurements between 7 and 10 a.m. were included, as well as several key climatic factors including wind factor, inversion factor, rain factor, and sun radiation.

Finally, the input file has ten input characteristics and one output (NO_2 day), which is the expected maximum NO_2 value for the rest of the day after 10 a.m.

5. Conducting the Experiment

We paid great attention in conserving the distribution of the 4 classes in the two sets, as it was in the initial data set. All algorithms were trained on the same training set, and were tested on the same testing set as well. The partitioning was a 70-30 partitioning. The same training and testing sets were used with all methods. At the end of the data preparation phase, we had one training set of 580 cases and one testing set consisting of 230 cases.

The training set was solely utilized for parameter tweaking and general data calibration during the NEMO system's training cycle. This included running the system repeatedly until the system parameters - the maximum and lowest number of prior cases retrieved, attribute weights, high and low attribute thresholds, similarity criteria, and adaptation technique - were fine-tuned and satisfactory results were produced.

We utilized cross validation testing on our training set to find the most efficient decision tree building technique among the axis split DT, the oblique split DT, and the CART-like DT. Six experiments were carried out, three of which used 5-fold cross validation and three of which used 10-fold cross validation. The CART method outperformed the other two in terms of accuracy, despite the fact that it built a bigger tree, which tended to over fit the input data.

After deciding to adopt the CART methodology for the decision tree experiment, and having already built the tree, we used the testing set, to justify the accuracy of the tree. A 10-fold cross validation tests, showed how well the tree represented the input file [21].

6. Results

The three algorithms for the training test produced comparable results. The testing set comprises 240 cases (30% of the whole data set), and the NO_2 concentration levels are distributed as follows:

Table 2. Classification of The Testing Set

Level	Cases	%
1	127	61
2	59	18
3	31	16
4	7	4

We're particularly interested in anticipating high NO_2 levels, thus predictive accuracy at levels 3 and 4 is preferable to that at levels 1 and 2. The CART-like strategy appears to yield superior outcomes for the two most crucial levels, based on the graphs above (3 and 4). Below is a graphical representation of the competitive accuracy:

The algorithms cannot clearly identify a first level instance from a second level scenario, which causes an obvious problem in forecasting the second level NO2 peaks by all three models. That means the triggering circumstances and associated characteristics are fairly similar in these two categories.

An actual predictive system requirement is clear indication whether current day is a NO2 episode day or not, giving this binary reply quickly and efficiently. According to the air pollution experts, this first decision is necessary to drive the subsequent phase of counter-measures proposal, i.e. issue of public warnings, traffic control measures, etc. For simplification purposes, a state reduction over the output is needed.

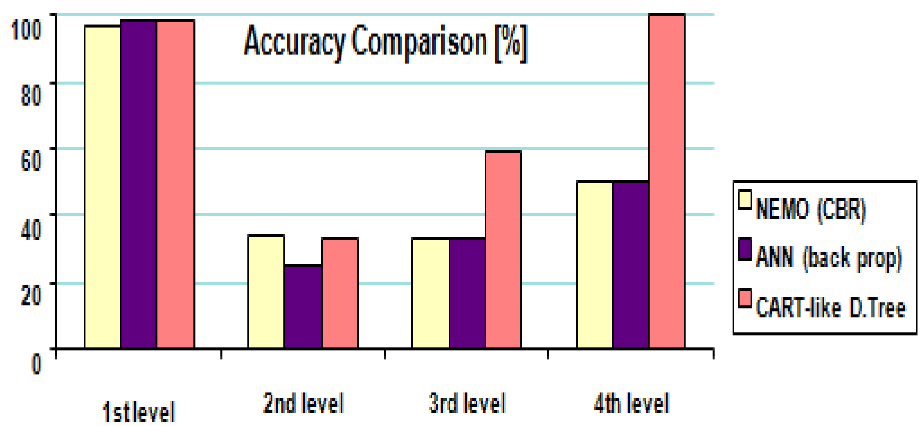

Fig. 4. Accuracy Comparison

For these reasons we merged the first and second level into one new level corresponding to no episode state, aswell as the third and fourth level into an episode state.

Using this as the success criterion, a new set of confusion matrices was built:

CART	No NO_2 episode predicted	NO_2 Episode predicted
No NO_2 episode	200	3
NO_2 Episode	12	25

ANN	No NO_2 episode predicted	NO_2 Episode predicted
No NO_2 episode	197	6
NO_2 Episode	16	21

NEMO	No NO_2 episode predicted	NO_2 Episode predicted
No NO_2 episode	200	3
NO_2 Episode	12	21

4 unpredictable cases

7. Discussion of System Performance

By interpreting the results of the conducted experiments, we can conclude that the three algorithms compete each other and obtain very close results. Overall, they constitute an efficient decision support ensemble that could function as a predictive working system in an air quality operational centre. Our findings indicate that CART classifiers outperform alternative tree-induction classifiers in complicated and multi-dimensional (spatial and temporal) domains of real-world applications, such as air quality prediction. It has also been demonstrated that the results produced using MLP neural networks are not as striking as in other areas, possibly due to high input data dependencies. Our NEMO CBR algorithm worked well when it came to predicting low NO2 emissions, but it struggled when it came to predicting large NO2 emissions.

All of the algorithms' outputs are not qualitative, although they can indicate whether a choice is binary or discrete. The algorithms' role is limited to that of a human expert decision support system since they have not been tested on a large historical database with numerous events, which would need many years of data that do not exist. Implementing a machine-learning system to evaluate the data collected from these sensors might lead to solutions for maintaining appropriate global temperature parameters, which have been steadily rising in recent decades. However, there is still a lot of work to be done. For large training sets, the suggested model performs better. However, training the model takes longer. The use of mobile sensors can aid in the development of a system to create an air pollution map. The accuracy of future estimates can be improved by incorporating meteorological parameters such as wind speed and weather into the algorithm.

8. Conclusion

We provide a novel machine learning-based approach for estimating dense air pollution utilising historical data from wireless sensor and government monitoring sites in this research. Atmospheric pollutants have a substantial impact on temperature, and these pollutants must be continuously monitored in order to develop temperature projections based on their quantity." Pollution sensor systems for data collection are now utilised to monitor the level of air pollution in all major cities across the world.

We may also evaluate the estimate performances of different Machine Learning models such as regression models (Decision Tree Regression, Random Forest Regression, and Linear Regression) to analyse the NO2 level. Predictions for additional pollutants such as CO, CO2, SO2, and others may be included in future studies. In the future, researchers will integrate the current findings with camera photos to analyse and anticipate air pollution in a variety of large cities. Future research aims include the creation of a platform that will provide solutions for traffic suggestions depend upon the sources of air pollution forecasts.

References

1. Department of Economic and Social Affairs: Urban Population Change; 2018. https://www.un.org/development/desa/en/news/population/2018-revision-of-world-urbanization-prospects.html. Accessed 20 Oct 2021.
2. Nada Osseiran, Christian Lindmeier: 9 out of 10 people worldwide breathe polluted air, but more countries are taking action; 2018. https://www.who.int/news/item/02-05-2018-9-out-of-10-people-worldwide-breathe-polluted-air-but-more-countries-are-taking-action Accessed 20 July 2021.
3. Ailshire JA, Crimmins EM. Fine particulate matter air pollution and cognitive function among older US adults. Am J Epidemiol 2014;180(4):359–66. https://doi.org/10.1093/aje/kwu155. https://academic.oup.com/aje/article-pdf/180/4/359/8640802/kwu155.pdf.
4. Pöschl U. Atmospheric aerosols: composition, transformation, climate and health effects. Angewandte Chemie Int Ed. 2005;44(46):7520–40. https://doi.org/10.1002/anie.200501122. Article Google Scholar
5. Du Y, Xu X, Chu M, Guo Y, Wang J. Air particulate matter and cardiovascular disease: the epidemiological, biomedical and clinical evidence. J Thoracic Dis. 2016;8(1):8.Google Scholar
6. Cohen AJ, Brauer M, Burnett R, Anderson HR, Frostad J, Estep K, Balakrishnan K, Brunekreef B, Dandona L, Dandona R, et al. 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 2015. Lancet. 2017;389(10082):1907–18. Article Google Scholar
7. Bu X, Xie Z, Liu J, Wei L, Wang X, Chen M, Ren H. Global pm2.5-attributable health burden from,. to 2017: estimates from the global burden of disease study 2017. Environ Res. 1990;2021(197):111123. Google Scholar OCDE.
8. The economic consequences of outdoor air pollution; 2016, p. 116. https://doi.org/10.1787/9789264257474-en. https://www.oecd-library.org/content/publication/9789264257474-en.
9. Mo Z, Huang J, Chen Z, Zhou B, Zhu K, Liu H, Mu Y, Zhang D, Wang S. Cause analysis of pm 2.5 pollution during the covid-19 lockdown in Nanning, China. Sci Rep. 2021;11(1):1–13. Article Google Scholar
10. Rodríguez-Urrego D, Rodríguez-Urrego L. Air quality during the covid-19: Pm2.5 analysis in the 50 most polluted capital cities in the world. Environ Pollut. 2020. https://doi.org/10.1016/j.envpol.2020.115042. Article Google Scholar
11. Zoran MA, Savastru RS, Savastru DM, Tautan MN. Assessing the relationship between surface levels of pm2.5 and pm10 particulate matter impact on covid-19 in Milan, Italy. Sci Tot Environ. 2020;738:139825. Article Google Scholar
12. Md N, Wai Y, Ibrahim N, Rashid Z, Mustafa N, Hamid H, Latif M, Er S, Yik L, Alhasa K, et al. Particulate matter (pm2.5) as a potential sars-cov-2 carrier; 2020.

13. Zhu Y, Xie J, Huang F, Cao L. Association between short-term exposure to air pollution and covid-19 infection: evidence from China. Sci Tot Environ. 2020;727:138704.Article Google Scholar
14. Wu X, Nethery RC, Sabath M, Braun D, Dominici F. Air pollution and covid-19 mortality in the United States: strengths and limitations of an ecological regression analysis. Sci Adv. 2020;6(45):4049. Article Google Scholar
15. Lu D, Mao W, Xiao W, Zhang L. Non-linear response of pm2.5 pollution to land use change in China. Remote Sens. 2021;13(9):1612. Article Google Scholar
16. Bartholomew DJ. Time series analysis forecasting and control. J Oper Res Soc. 1971;22(2):199–201. https://doi.org/10.1057/jors.1971.52.Article Google Scholar
17. Kumar U, Jain V. Arima forecasting of ambient air pollutants (o 3, no, no 2 and co). Stochastic Environ Res Risk Assess. 2010;24(5):751–60.Article Google Scholar
18. Yu R, Yang Y, Yang L, Han G, Move OA. Raq-a random forest approach for predicting air quality in urban sensing systems. Sensors. 2016;16(1):86.Article Google Scholar
19. Lin K-P, Pai P-F, Yang S-L. Forecasting concentrations of air pollutants by logarithm support vector regression with immune algorithms. Appl Math Comput. 2011;217(12):5318–27.MATH Google Scholar
20. Wang P, Liu Y, Qin Z, Zhang G. A novel hybrid forecasting model for pm10 and so2 daily concentrations. Sci Tot Environ. 2015;505:1202–12.Article Google Scholar
21. Mishra D, Goyal P. Neuro-fuzzy approach to forecasting ozone episodes over the urban area of Delhi, India. Environ Technol Innov. 2016;5:83–94.Article Google Scholar
22. Zaidan MA, Dada L, Alghamdi MA, Al-Jeelani H, Lihavainen H, Hyvärinen A, Hussein T. Mutual information input selector and probabilistic machine learning utilisation for air pollution proxies. Appl Sci. 2019;9(20):4475.Article Google Scholar
23. Danesh Yazdi M, Kuang Z, Dimakopoulou K, Barratt B, Suel E, Amini H, Lyapustin A, Katsouyanni K, Schwartz J. Predicting fine particulate matter (pm2. 5) in the greater London area: an ensemble approach using machine learning methods. Remote Sens. 2020;12(6):914.Article Google Scholar
24. Salman AG, Heryadi Y, Abdurahman E, Suparta W. Single layer & multi-layer long short-term memory (lstm) model with intermediate variables for weather forecasting. Procedia Comput Sci. 2018;135:89–98.Article Google Scholar
25. Tsai Y-T, Zeng Y-R, Chang Y-S. Air pollution forecasting using rnn with lstm. In: 2018 IEEE 16th Intl Conf on dependable, autonomic and secure computing, 16th intl conf on pervasive intelligence and computing, 4th intl conf on big data intelligence and computing and cyber science and technology congress (DASC/PiCom/DataCom/CyberSciTech). IEEE; 2018, p. 1074–9.
26. Shi P, Zhang G, Kong F, Chen D, Azorin-Molina C, Guijarro JA. Variability of winter haze over the Beijing-Tianjin-Hebei region tied to wind speed in the lower troposphere and particulate sources. Atmos Res. 2019;215:1–11.Article Google Scholar
27. Pohjola MA, Kousa A, Kukkonen J, Härkönen J, Karppinen A, Aarnio P, Koskentalo T. The spatial and temporal variation of measured urban pm 10 and pm 2.5 in the Helsinki metropolitan area. Water Air Soil Pollut Focus. 2002;2(5):189–201.Article Google Scholar
28. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521(7553):436–44.Article Google Scholar
29. Garg S, Jindal H. Evaluation of time series forecasting models for estimation of pm2.5 levels in air. In: 2021 6th international conference for convergence in technology (I2CT). IEEE; 2021, p. 1–8.
30. Xayasouk T, Lee H, Lee G. Air pollution prediction using long short-term memory (lstm) and deep autoencoder (dae) models. Sustainability. 2020;12(6):2570.Article Google Scholar
31. Jeya S, Sankari L. Air pollution prediction by deep learning model. In: 2020 4th international conference on intelligent computing and control systems (ICICCS). IEEE; 2020, p. 736–41.