Journal of Chemical Health Risks

www.jchr.org

JCHR (2023) 13(5), 278-301 | ISSN:2251-6727

Air Pollution Surveillance Systems: A Review of the modelling and the Forecasting Technologies

[1,*]Bhupesh Patra, [2]Abha Tamrakar

[1]Research Scholar of School of Information and Technology, ISBM University, Gariyaband, Chhattisgarh, India

[2]Associate Professor of School of Information and Technology, ISBM University, Gariyaband, Chhattisgarh, India

[*]Corresponding author email-bhupeshpatra0@gmail.com

(Received: 02 September 2023 Revised: 14 October Accepted: 07 November)

KEYWORDS

Air pollution surveillance system; Big data technologies; Sensors; Machine learning;IoT.

ABSTRACT:

Air pollution is a major concern, particularly in modern cities due to its significant effects on public health and the global economy. The importance of the air quality information makes the need for extremely precise real-time monitoring systems essential. The researchers are compelled to create future air pollution surveillance systems by utilising cutting-edge technologies like machine learning techniques, big data technologies, sensors, and the Internet of Things for suggesting a stable and effective model for the stated purpose due to the conventional air monitoring systems' limited data access, high cost, and inability to be scaled up. The ability of machine learning algorithms to forecast air pollution by means of general pattern and abrupt changes is demonstrated by extensive real-time air pollution trials. For gathering and analysing air data, smart devices are required. By analysing and gathering recent research in this area, this review paper focuses on providing an overview of air pollution surveillance systems (APSS) and emphasises data sources, monitoring, and forecasting models to enhance the various components of air polluting models. Additionally, it provides light on a variety of research-related problems and difficulties.

1. INTRODUCTION

Modern human activities inevitably involve energy usage and its effects. The combustion of the straw, coal, and kerosene are only a few examples of the human-caused causes of air pollution, along with emissions from factories, cars, planes, and aerosol cans. Every day, numerous harmful pollutants such as CO, CO_2, Particulate Matter (PM), NO_2, SO_2, O_3, NH_3, etc. are emitted into our environment. The health of people, animals, as well as plants is impacted by the chemicals and particles that make up air pollution. Humans are susceptible to a wide range of dangerous illnesses brought on by air pollution, including pneumonia, lung cancer, heart disease, and bronchitis. Smog, aerosol production, impaired vision, rising temperatures, acid rain, early mortality, and other current environmental problems are all caused by poor air quality. The atmospheric emissions from factories and power plants, as well as exhausts from agricultural and other sources, are to blame for the rise in greenhouse gases. The greenhouse gases have a negative impact on the climate, which in turn affects plant development. Plant-soil interactions are impacted by discharges of elemental carbons and greenhouse gases as well[1,2]. Climate change has a significant impact on agricultural aspects and production in addition to humans and other animals[3]. A measurement metric known as the Air Quality Index (AQI) has a direct connection to public health. A higher AQI level denotes a riskier exposure for the general public. Therefore, the desire to accurately anticipate the AQI drove scientists to track and model air quality. With

Journal of Chemical Health Risks
www.jchr.org
JCHR (2023) 13(5), 278-301 | ISSN:2251-6727

an increase in industrial and motorised growth, monitoring and forecasting AQI, particularly in metropolitan areas, has become an essential and difficult undertaking. Although the concentration of the deadliest pollutant, PM2.5, is shown to be multiplied in developing countries [4].

A few researchers attempted to do the study of Indian city air quality prediction. After reading the available research, it became clear that there was a great need to close this gap by undertaking analysis and AQI prediction for India[5].

Air pollution has reached dangerous levels as a result of increased transportation, population growth, rising global temperature, and sudden climate changes. In order to create a safer and better environment for people, animals, and plant life, pollution must be monitored and controlled. Conservation authorities and the government have made fantastic efforts to lessen the effects of air pollution on the community. Researchers, decision-makers, and developers can control and enhance the living environment with the use of comprehensive data regarding the air contamination state. Typically, the traditional air pollution monitoring stations evaluate the quality of the air. These stations for monitoring have the ability of measuring a wide range of contaminants and have good data accuracy. Researchers have recently focused on the alternative solutions and given them the name of a potential air quality surveillance system due to the incomplete data access, big size, high cost, and non-scalability of the current air monitoring stations[6,7].

These systems make use of cutting-edge sensing technologies such as IoT, WSN, and inexpensive ambient sensors. These systems are still in their infancy and have a number of problems, including a lack of 3-dimensional data acquisition, compatibility problems, decreased accuracy, and network scalability[8]. The monitoring of air pollution has received a lot of attention recently, but neither the existing nor the suggested methods are able to give precise and cost-effective geographical and temporal resolutions of the pollutant information. The majority of IoT-based monitoring systems statistically show the measured number for each pollutant, making it challenging for regular users to determine the amount of air pollution. For measuring the various contaminants in a variety of environments, the standing, portable air surveillance devices are insufficient[9].The advanced technologies such as IoT, Big Data, and ML-based Air Pollution Monitoring and Prediction Systems are shown in Figure 1.

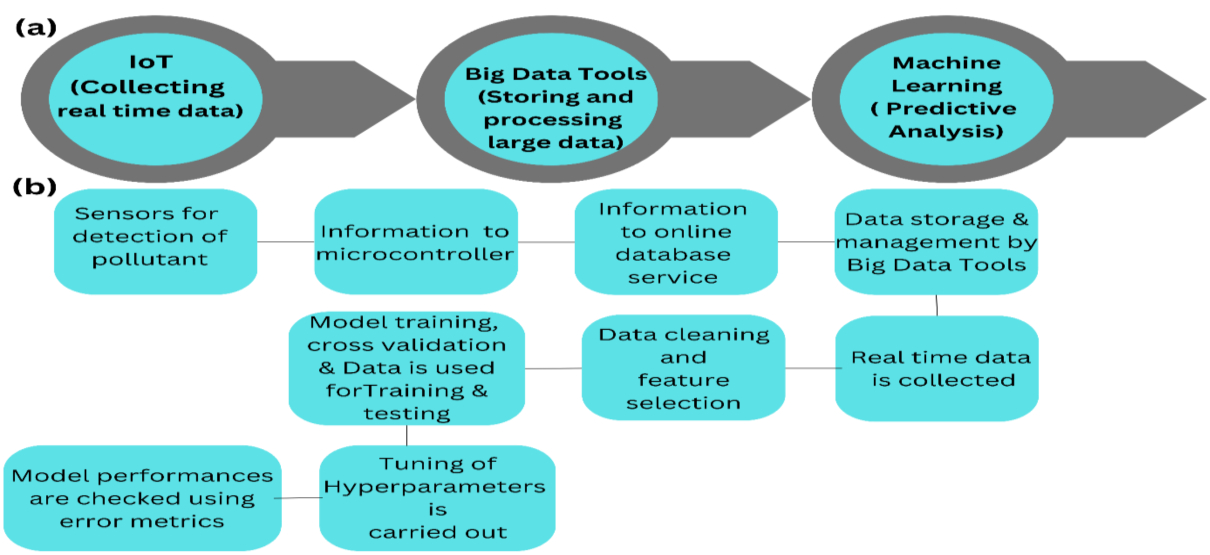

Figure 1: Air pollution monitoring and prediction through IoT, Big Data, and ML-based Systems

Journal of Chemical Health Risks
www.jchr.org
JCHR (2023) 13(5), 278-301 | ISSN:2251-6727

The literature has used a variety of models to predict AQI, including mathematical, deterministic, physical, and machine learning (ML) models. Traditional methods that rely on statistics and probability are extremely complicated and ineffective. The models based on machine learning for AQI prediction have shown to be more consistent and dependable. Data collecting is now simple and precise thanks to advanced technology and sensors[10].Only machine learning (ML) algorithms can effectively handle the rigorous analysis needed to make accurate and trustworthy predictions from such vast environmental data. This article provides a comprehensive overview of the most recent air quality monitoring technologies, including Machine Learning (ML) simulations, inexpensive sensors, and air surveillance system deployment methods. The major design criteria for air pollution surveillance systems are addressed as well, along with upcoming research difficulties.

2. AIR QUALITY AND ITS MONITORING

We inhale millions of times each day without giving it any thought until we start to feel uncomfortable. The most dangerous type of pollution, air pollution, is caused by a variety of things, including human activity and industrial processes that emit solid, liquid, and gaseous chemicals into the environment on a daily basis.

The World Health Organisation (WHO) has recently released statistics showing that air pollution causes the deaths of about 7 million people year. Three million people die prematurely each year as a result of air pollution; these deaths are caused by exposure to minute particles with a diameter of 10 microns or less, which can cause cancer, heart disease, and lung illness. Furthermore, if the sources of air pollution are not identified, it is extremely challenging to address them. A monitoring system for air quality can assist the public keep an eye on air pollution. These monitoring systems provide regular monitoring of the air quality[11]. The public can be guided by air monitoring to take the proper decisions and activities in accordance with each person's health requirements. Additionally, it raises public awareness of pollution, which may result in the outlawing of activities that contribute to pollution on a large scale.

2.1 Air Pollution

Air pollution can occur indoors or outside and refers to the degradation of the air. It happens when airborne contaminants infiltrate the atmosphere and contaminate the air, making it impossible for humans, animals, and plants to exist. All living creatures are supported by a collection of gases that together make up the atmosphere. It may be dangerous to survive due to the imbalance brought on by changes in these gases[12].

2.1.1 Air Quality

The quality of the air in our immediate environment is referred to as air quality. The extent that the air is clear, clean, and free of pollutants like smoke, dust, and other types of gas contaminants in the air is referred to as its air quality. The intricate equilibrium of life on earth, which includes people, creatures of all kinds, and resources from nature, must be preserved, and is put into danger when pollution levels in the air approach dangerously high levels[13].

2.1.2 Effect of pollutants on Air Quality

Air quality is continuously being degraded by emissions from many sources. These come from either natural or artificial sources. Volcanic eruptions, windstorms, biological deterioration, and forest fires are a few examples of natural causes. Pollution from moving automobiles, production sites, power stations, furnaces, and burning coal or wood products are examples of human-made sources. These sources emit pollutants into the air, which can have a serious negative impact on people, animals, and the ecosystem. The quantity of pollutants, their rate of atmospheric release, and the length of time they are confined in a region are the three variables that determine the quality of the air we breathe. Pollutants in the air are likely to interact with the air and swiftly disperse if they are in a region with good airflow. When there are particular circumstances, such as light winds or impediments that prevent the passage of these harmful substances away from a location, air pollutants have a tendency to stay in the air. As a result, the concentration of air pollution rises quickly[12].

Journal of Chemical Health Risks

www.jchr.org

JCHR (2023) 13(5), 278-301 | ISSN:2251-6727

2.2. Types of the Air Pollutants

The two types of air pollutants that are often distinguished are primary pollutants and secondary pollutants. Natural or human-induced activity increases primary air pollutants, but interactions between primary pollutants produce secondary pollutants[14].

Smog is brought on by the mixing of a number of main pollutants, while sulphur dioxide is created by burning fossil fuels in industrial facilities and by automobiles. The primary causes of outdoor air pollution are the use of fossil fuels in manufacturing processes, transportation, agriculture, and mines.

3. AIR POLLUTION MONITORING

With regard to the location and the environment, air pollution surveillance systems (APSS) have the potential to be divided into three primary groups, such as indoor, outdoor, and industrial pollution monitoring.

Indoor air pollution occurs in small enclosed spaces such as houses, workplaces, offices, and closed spaces like subterranean shopping malls and tube stations[15]. Air pollution from the outdoors corresponds to the open environment, which includes every aspect of the atmosphere and everything else. The industrial environment is a particularly delicate place where harmful gases and other substances are present. Monitoring systems for indoor, outdoor, and industrial air have unique associated requirements due to their various environments and pollutant types, as detailed below.

3.1. Indoor APSS

Despite the fact that the types of contaminants are different, unexpectedly enclosed places have worse air quality than open areas, probably because of inadequate air circulation and inadequate ventilation. The goods utilised in household and workplace activities as well as these activities themselves significantly contribute to environmental contamination. The oxides of nitrogen, sulphur oxides (SO_2), carbon monoxide (CO), carbon dioxide (CO_2), methane (CH_4), volatile and semi-volatile organic compounds, particulate matter, relative humidity, and microorganisms are among the indoor air hazards recognised by experts. All or the majority of these significant pollutants should be detectable by the system created for indoor monitoring. For such systems, electricity is not a huge concern because they can be powered on directly and networks may be kept running as long as necessary. Environmental and atmospheric elements have little or no influence on indoor pollutants. Since the capacity to recognise pollution patterns is sufficient for the user to be alerted and take appropriate action, these systems only need moderate data accuracy. The fundamental objective of such systems is public monitoring, and the average person should be able to keep an eye on the air quality in the vicinity of his or her living areas. It should be simple to deploy and maintain, and it should be user-friendly[16–18].

3.2. Outdoor APSS

Burning fossil fuels, manufacturing waste, industrial activities, transportation systems, agricultural practises, and mining processes are the main factors that cause outdoor air pollution. NO_x, SO_x, carbon monoxide, and particulate matter (PM2.5 & PM10) should all be able to be measured by an outdoor air pollution monitoring system.

Environmental and climatic parameters like temperature, humidity, altitude, wind speed, and pressure have a significant impact on outdoor pollutants. For reliable measurement, the system of surveillance should be able to counteract all of these variables. The dispersion of pollution in the outdoor environment is not uniform and varies on the exact location of the source of pollution, therefore the locations of the monitoring system's nodes need to be carefully chosen in order for it to function properly. Urban pollution is closely correlated with societal behaviour and monitoring sites. The need for coverage in every dimension poses formidable difficulties for outdoor monitoring. Multiple sensing nodes are needed to cover a sizable geographic region over the whole location, and because these nodes are battery-operated, the network life and power consumption are the systems' bottlenecks. Deployment and maintenance shouldn't be very difficult or time-consuming[19–21].

Journal of Chemical Health Risks
www.jchr.org
JCHR (2023) 13(5), 278-301 | ISSN:2251-6727

3.3 Industrial APSS

Industries are more vulnerable to being exposed to dangerous contaminants. Inhaling harmful substances and gases carries a very significant risk. The APSS's development in the manufacturing sector is of the utmost importance. The type of industry and surrounding environment have an impact on the pollutants' nature. Of the systems mentioned above, these systems have sporadic requirements because of the great sensitivity of this class and the need for quick responses, excellent accuracy.Cost and power consumption are secondary concerns for this particular type of APSS, thus implementation and maintenance should be comparatively simple[22].

4. POLLUTION MONITORING SYSTEM OVERVIEW

Currently, the traditional monitoring stations are used to monitor air pollution. These stations provide excellent consistency, accuracy, and the ability to count a wide range of contaminants using analytical tools including gas chromatograph-mass spectrometers. The main objective and basic design of the existing monitoring systems are the same, even though they employ different technology concepts and tactics (Figure 2).

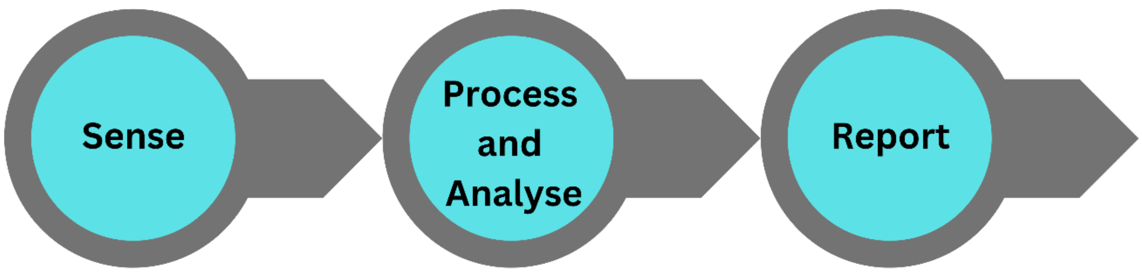

Figure 2 Fundamental building blocks system for an air monitoring

The three fundamental building components of an air monitoring system are typically sense, process, and report. The sense block collects the necessary information from the environment, the process & analyses block is in charge of gathering and analysing the sensed data and producing the corresponding results, and the report block is in charge of visualising, storing, and issuing warnings. For various scenarios, sensing equipment's processing level, hardware, and software tools vary, but it depends on the application needs, such as the viability and accessibility of the tools for prototype or implementation[23,24].

4.1. Regulatory Bodies & Standards

4.1.1 Air Quality Standards

The air quality indicator (AQI) is a pointer for pollution level presented by environmental protection authorities for better understanding. The air quality index (AQI) describes how clean the air is now or will be in the future. The definition of an index differs from one country or region to another based on the air quality catalogues used. Different standard limits for six major pollutants have been set by the United States Environmental Protection Agency (EPA), the World Health Organisation (WHO), the European Commission (EC), the Chinese Ministry of Environmental Protection (MEP), and the Environmental Protecting Department (EPD) of Hong Kong (see next subsection). Two metrics (ATMO & IQA index) are standardised in France to improve air quality observation[25].

4.1.2 Air Quality Index AQI

The air quality index, or AQI, is a benchmark for measuring the level of air pollution and providing a quantitative interpretation. According to the EPA and MEP, the major pollutants that affected the analysis are minute particulate matter (PM2.5), inhalable particulate matter (PM10), SO_x, NO_x, and CO. CO is scaled in parts per million (ppm), SO_x

Journal of Chemical Health Risks

www.jchr.org

JCHR (2023) 13(5), 278-301 | ISSN:2251-6727

and NO_x are scaled in parts per billion (ppb), and PM2.5 and PM10 are scaled in micrograms per cubic metre (g/m3)[25].

4.2 Hardware & Software Tools

Different platforms, pieces of hardware, and software are employed by the researchers. STM32F103RC was used to deploy the monitoring node. Gizduino, waspmote, and mega128 were also used as sensing nodes. IoT gateways commonly employ Arduino, GSM modules, raspberry pi 3, and USRP as an IoT gateway node. Lab View (for user applications), Mat Lab, Android Apps, MYSQL (database management system), XML, XAMPP, Apache (HTTP server), DBMS (database-management), and Thing Speak are examples of software tools for user application development and data administration[25].

4.3. Communication Protocols

The APMS uses two different categories of communication protocols. The first type of communication is short-distance (between sensing nodes), while the second type is long-distance (between sink nodes and the IoT gateway). Zigbee (short-range, between sensors), Bluetooth, WIFI, GPRS (IPV4, IPV6), LTE Network, GSM (General System Mobile), and radio frequency are just a few of these communication protocols[25].

4.4. Low-Cost Pollution Sensors

There are many different types of pollution sensors on the market, primarily gas detectors and particle matter sensors. The subsections below compare all the available sensors in terms of their ability to detect various gases, cross sensitivity, cost, lifespan, and power consumption. Figure 3 depicts the sensors used for air surveillance systems.

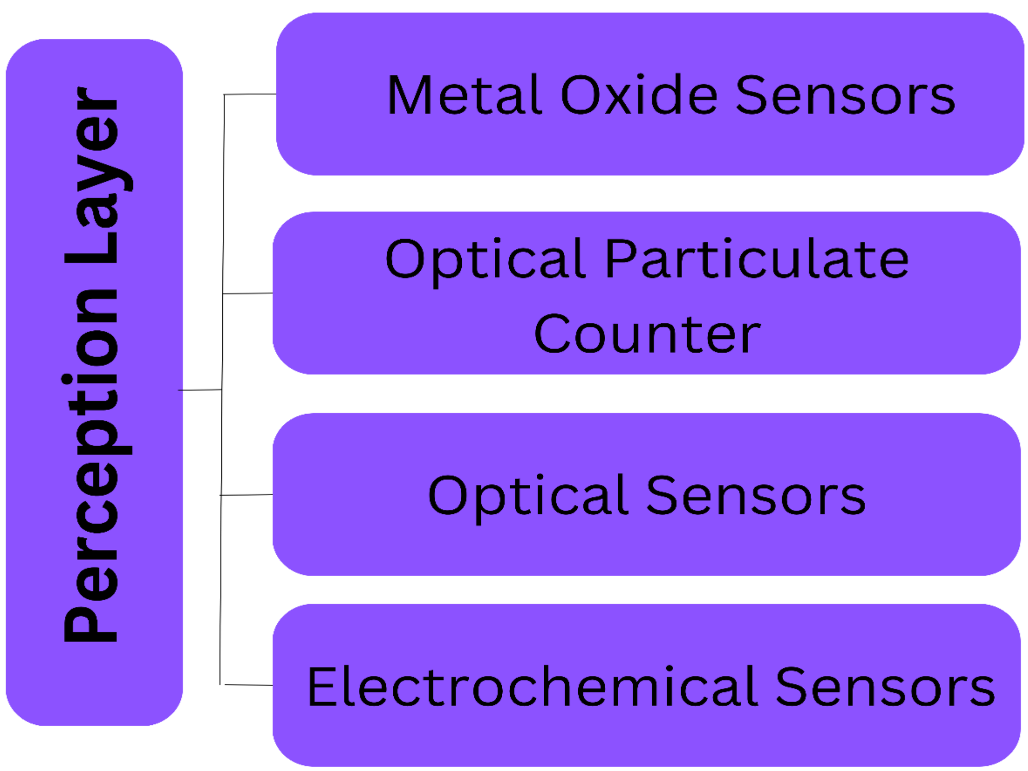

Figure 3 Different types of sensors for air surveillance system

4.4.1. Gas Sensor

Many advancements in the field of gas sensors have been made recently, but each technology utilised has pros and downsides of its own. The most common types of gas sensors include electrochemical (EC), catalytic, solid-state, photo-ionized, and non-dispersive infrared radiation absorption (infrared absorption). These sensors are inexpensive, compact, and respond quickly (from a few seconds to a few minutes).

These sensors fall into various categories, each of which can identify a particular class of gases. EC sensors typically have two electrodes, but modern EC sensors incorporate one or two additional electrodes to increase stability and sensitivity. Considering the measurement procedures, EC sensors can be divided into three primary categories: conductometric, amperometric, and potentiometric[26]. Amperometric gas sensors have a stellar reputation for reliability in industrial air monitoring and safety. They benefit from a signal current generated which is proportional to the measured gas concentration. A sensing layer is present in metal oxide sensors, and it is utilised to either absorb or desorb the target gas. This layer needs to be heated to 250℃ in order to increase its sensitivity.

These inexpensive sensors can give adequate data accuracy and a respectable detection range, but they still fall short of the accuracy attained by traditional monitoring stations[27]. These sensors also have the downside of requiring pre-calibration before use and recalibration after a set amount of time. The calibration process requires subjecting the detector to the desired gas (that has to be detected) at a specified concentration. In order to obtain accurate results, various sensor parameters are then modified. Carbon monoxide (CO), ozone (O_3), sulphur dioxide, and nitrogen dioxide are the gases that need to be monitored for the three different types of air pollution.

4.4.2 Particulate Matter Detectors

The mass concentrations of particulate matter (PM) can be measured using a variety of methods, but this is not an easy task. PM is challenging by nature; different measurement approaches may produce inconsistent results.

To regulate the fluctuating temperature and humidity, a heating element was used. However, because it evaporates the semi-volatile species, it alters the measurement findings. To counteract these effects, some setups utilise specialised dryers. There are two primary categories of PM measurement methods. One type of gravimetric sampler is direct reading, while the other is filter-based.Measurements of the amount of substance in ambient air are provided by direct reading. The lab process for collecting PM and weight using filter-based approaches takes a lot of time and resources[25,28,29].

4.4.3 Performance Limiting Factors and Error Sources

The greatest obstacle to the reliable deployment of cheap sensors in monitoring systems is their low measurement accuracy. After years of sensors evaluation and signal analysis, researchers have discovered a large number of error causes, which they divide into two primary categories: internal mistakes and external errors[30]. Internal faults, such as dynamic boundaries, systematic flaws, asymmetric responses, and signal drift, are connected to the operation of low-cost sensors. Environmental factors are the main cause of external error sources, which heavily depend on the model and the deployment area. Low selectivity and environmental dependency are the two primary external error drivers. By using a model that associates the measurements of a low-cost sensor with an accurate reference, calibration is carried out to lower the inaccuracies of low-cost pollution sensors. Both prior to and after the sensors are installed, calibration of the sensors is required for optimal performance. This can be accomplished using a variety of mathematical techniques, and multiple calibration models exist for various types of sensors. Errors resulting from dynamic boundaries and systematic flaws are corrected by offset and gain calibration. Measurements are extended by temperature and humidity correction using current readings of temperature and humidity.

Interfering gases from the environment are eliminated via sensor array calibration. Low-cost sensors need to be periodically recalibrated in addition to pre-calibration. Some network-based calibrations include transfer calibration, collaborative calibration, and blind calibration[28,30,31].

5. MONITORING SYSTEM DEPLOYMENT STRATEGIES

As a result of their complexity and need for extra gear, conventional monitoring stations are expensive to maintain. Air pollution algorithms or approximations monitor the areas where the stations for monitoring are not installed, but this data lacks cross-validation and confirmation. Researchers are implementing a novel way for monitoring systems that combines the WSN, IoT, and low-cost sensors into a single system. Additionally, these sensors provide the flexibility and practicality of using the sensor nodes on a broad scale. These WSN-based models assist the researchers in accurately and competently diagnosing the spread of pollutants to enhance the prototypes. There are many different methods for monitoring the quality of the air, and a recent study has divided them into two broad groups: stationary air monitoring systems and portable air monitoring systems[32].

5.1. Stationary APSS

Sensing nodes are used in stationary surveillance systems, which are set up in fixed places. The number of sensing nodes relies on the model, which is covered in more detail in the subsections following.

5.1.1. Traditional Monitoring Stations

The majority of traditional air quality monitoring systems are made up of elegant, time-tested equipment. The precision of the data is increased by these systems' use of sophisticated measuring algorithms and auxiliary equipment such temperature controllers, relative humidity controllers, air filters, and calibrators[11]. Because of their high cost, high power consumption, large size, and weight, these monitoring stations are typically impossible to be deployed across a large region.

5.1.2. Portable Monitoring Devices

Recent research has used a variety of techniques to create portable monitoring systems, especiallyhaving one sensing node with numerous sensors. To track indoor/outdoor air pollution, single sensing nodes are developed. Though the precision is not particularly great, these portable instruments can only test a few pollutants, but they can nevertheless give an estimate of the individual's air exposure[31,33].

5.1.3. Wireless Sensor Network Based Approaches

These techniques adapt the WSN approach to gather data, process sensed data, and then deliver it to the user. The general public can access accredited data via websites, mobile apps, etc. Static modules of sensors are frequently mounted on streetlight shafts or other carefully chosen locations. These systems have many advantages over traditional monitoring stations, including looser energy consumption restrictions (deployment dependant), the elimination of the requirement for locating modules like GPS because the nodes are fixed, guaranteed network connectivity, and ease of maintenance. Along with the aforementioned benefits, the system has a few downsides, such as the need for a custom-built network, inconvenient calibration & sensor drift (cheap cost - low quality), and two-dimensional data measurement[34].

5.1.4. IoT based Approaches

The development of the inexpensive APSS is made possible by sensor and IoT technologies. IoT-based monitoring systems that use cheap sensors and a variety of communication methods. These devices are designed to detect the ambient air quality in actual time and allow users to respond appropriately. The system's primary function is to alert users about the current air quality ranking more subtly and in real time. The application of IoT to environmental safety introduces a form of real-time air quality monitoring. This Internet of Things (IoT) base system can save hardware costs on a high base and is deployable over a wide area to provide high special coverage. IoT technology is used to create the monitoring system because an IoT system can have problems with compatibility because it consists of disparate devices from various vendors[35,36].

5.2. Mobile APSS

Sensing nodes are placed on moving bodies, which can be people, shared vehicles, or special vehicles, in mobile APSS. Such systems can have one or more sensing nodes, although there are fewer of them than in a static example. Two types of mobile monitoring systems emerge when used in a WSN fashion: the municipal sensor network (MSN) and the automobile sensor network (ASN)[37].

5.2.1. Vehicle as a Carrier/ASN

The sensing modules in ASN systems are often installed on vans or other specifically equipped vehicles used for public transportation. The mobility of cars enables the sensor module to cover the necessary wide geographic area. The general public can access accredited data through websites and mobile apps, however ASN is not economical for carriers.

When using public transit as a carrier, carrier speed is unpredictable and cannot be altered in accordance with needs, which results in inaccurate data[38].

5.2.2. Community/Public Supported/MSN

In this method, the public societies, generally by unpaid assistance, govern and maintain the sensor nodes. Operators have the ability to collect, analyse, and disseminate local air quality data. Although this system lacks centralised management and is very cost-effective, the data it yields is a public asset and is not very authentic or dependable[39].

5.3. Trade-off between the existing strategies

When dealing with these previously planned monitoring systems, it was evident that there were notable trade-offs. For conventional monitoring stations (CMS), static sensor networks (SSN), municipal sensor networks (MSN), and automotive sensor networks (ASN), Figure 4 illustrates the trade-off among cost, spatial coverage, and the likely data accuracy. The results show that the price tends to go up as you move through top to bottom, but that the tendency for geographical coverage is the opposite of that of cost and accuracy[25].

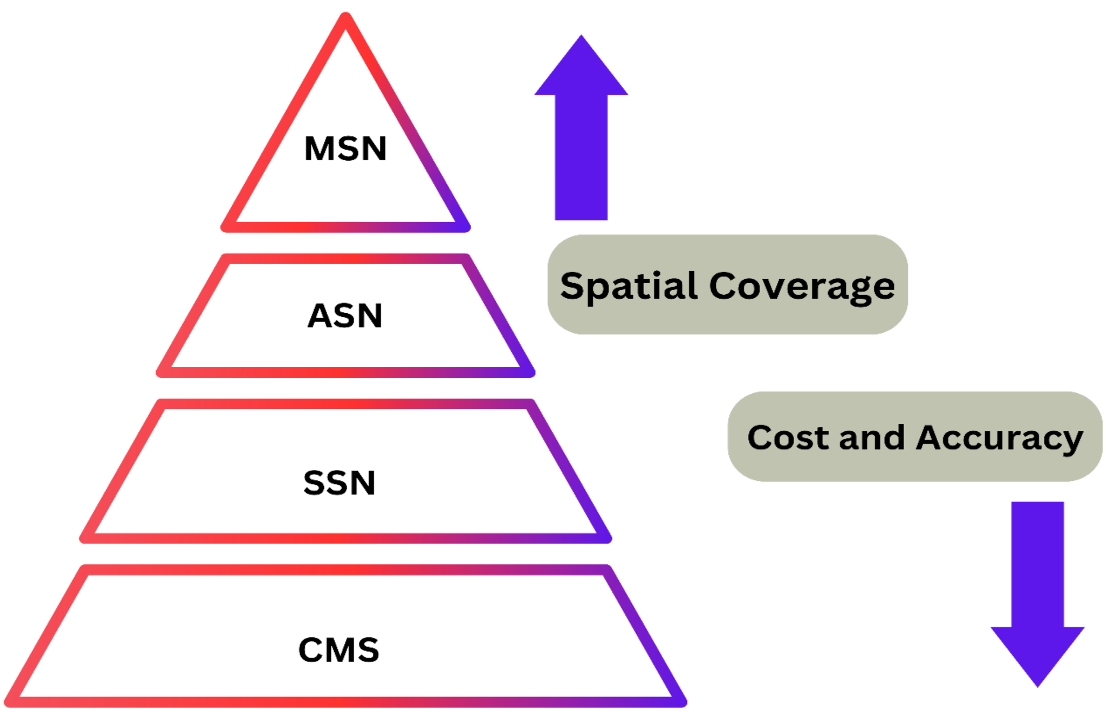

Figure 4 Trade-off among cost, spatial coverage, and the likely data accuracy for conventional monitoring stations (CMS), static sensor networks (SSN), municipal sensor networks (MSN), and automotive sensor networks (ASN),

Journal of Chemical Health Risks
www.jchr.org
JCHR (2023) 13(5), 278-301 | ISSN:2251-6727

6. AIR MONITORING SYSTEM ARCHITECTURES

After discussing the many types of air pollution surveillance systems, this section will quickly go over each system's architecture.

6.1 WSN based Architecture

A wireless sensor network (WSN) is made up of a large number of sensing nodes that communicate wirelessly with one another. WSNs have a lot of promise for widespread usage in the monitoring fields. However, only a small number of WSNs have had their potential for complete utilization of monitoring of the air pollution.

Figure 5 depicts the architecture of WSN-based air monitoring systems. End-user devices (source /sensors) communicate with one other and with sink nodes (coordinators) through a mesh network. Collected data is sent to the middleware via sink nodes, where additional processing is carried out before being delivered to the end user. The example makes use of the mesh topology even though alternative topologies including star and tree are also used in some of the described systems[25].

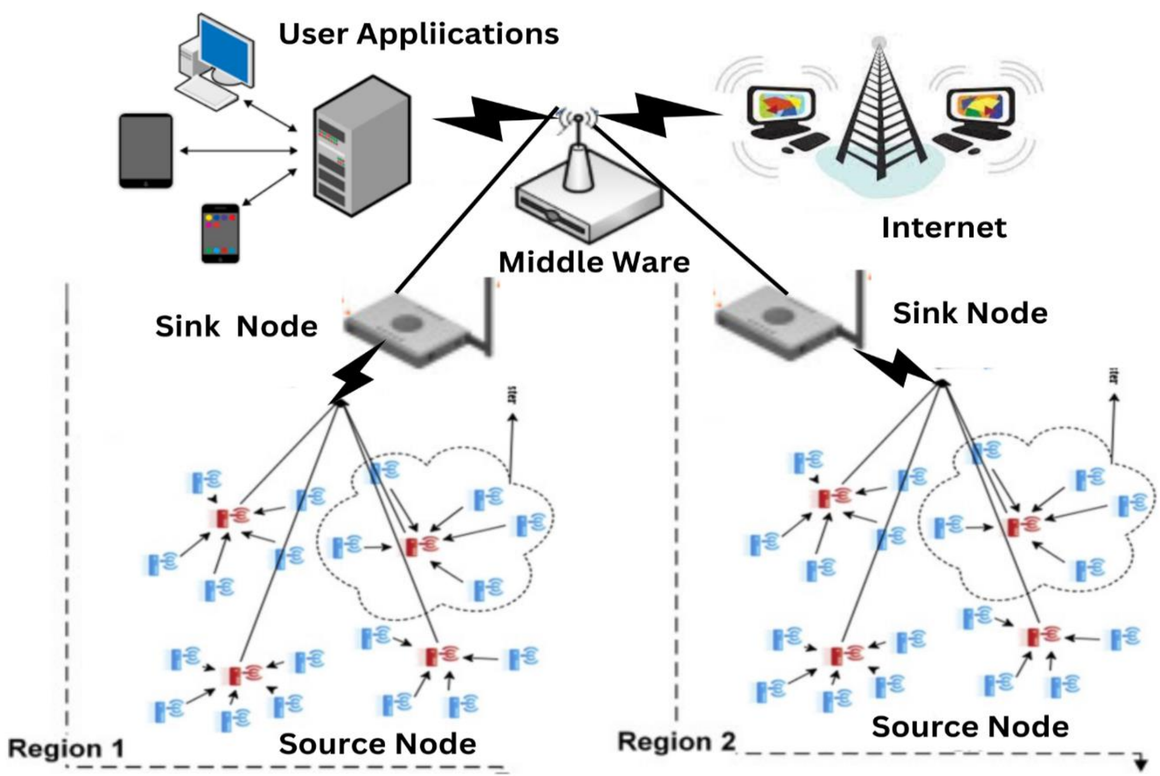

Figure 5 Architecture of WSN-based air monitoring systems

6.2. IoT based Architecture

The sensors used in an IoT-based monitoring system gather data on air quality in real time, and the IoT cloud is used to store and analyse all of the data.According to IoT architecture, the system is made up mostly of the application layer, the network layer, and the sensor layer. The basic layout of the monitoring system is depicted in Figure 6, which also includes the protocols used on each layer by the various system developers. Each system may have a little

Journal of Chemical Health Risks
www.jchr.org
JCHR (2023) 13(5), 278-301 | ISSN:2251-6727

difference in architecture depending on the requirements. WSN technology may be used at the sensing layer, which houses the monitoring nodes. For instance, field nodes can be deployed quickly and with high reliability, they can be remotely configured and updated, they are resilient to long-term environmental exposure, they can send out programmed multichannel warnings, and they can last for three to ten years without requiring maintenance[21,40].

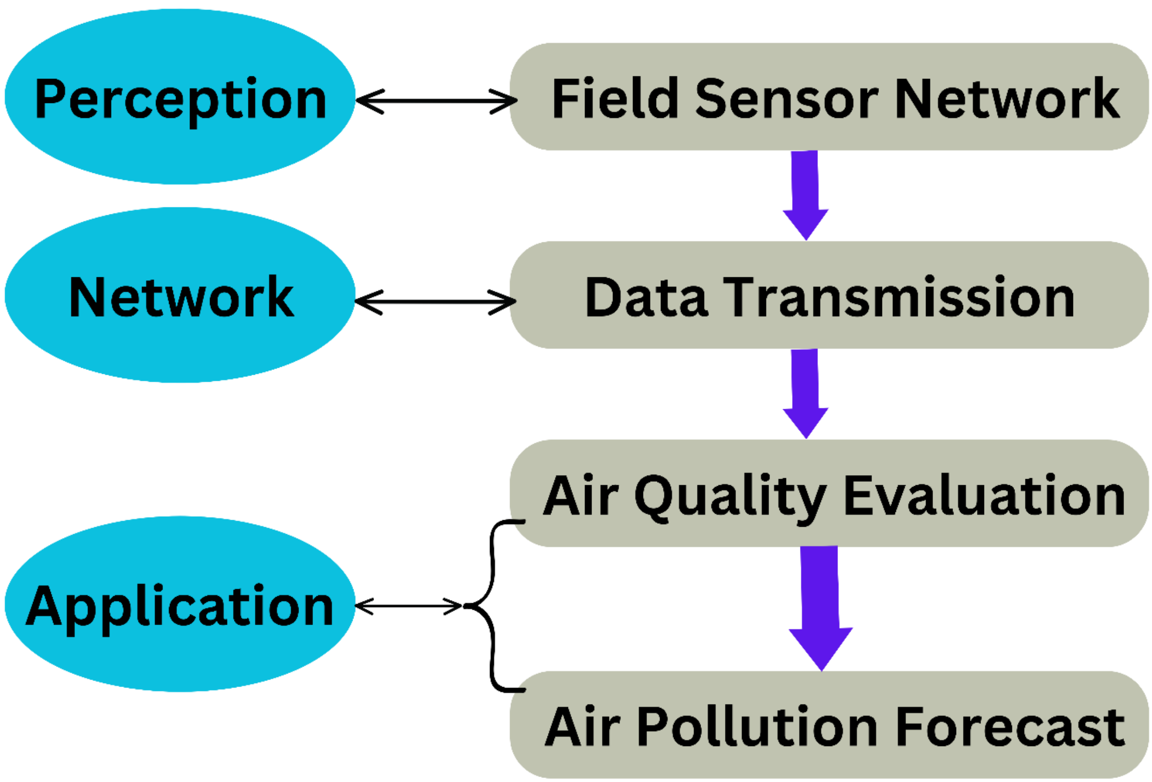

Figure 6 IoT-based air monitoring system

7. OPEN RESEARCH CONCERNS AND CHALLENGES

The current investigation covered monitoring systems connected to inexpensive sensors. Gaining a high degree of accuracy in these systems with low-cost sensors is a significant challenge[21]. These low-cost sensors have the ability to offer adequate enough data accuracy and reasonable detection range, but they are still unable to compete with the level of precision obtained from the conventional monitoring stations. How data depart from the original reference depends on the degree of data uncertainty. Raw data from inexpensive sensors is further processed to raise the quality to a level where a reference is required. It is quite challenging to get a true reference in the deployment environment because there is no control over the situations in real environments. The co-location of reference sensors with every low-cost sensor in the network would be almost impossible to manage. As a result, the correct calibration of

Journal of Chemical Health Risks

www.jchr.org

JCHR (2023) 13(5), 278-301 | ISSN:2251-6727

networks remains an issue that could be partially overcome by carefully considering the factors listed below[34].

i) Identification and analysis of the ambient factors that primarily influence the sensor's response.

ii) Examining the presence and root causes of anomalies in environmental data.

iii) Ways to have more dependable and varied references for assessing cheap sensors.

Community-supported air monitoring systems use volunteers or public transport to carry sensors. In these systems, the carriers are unregulated or just partially controlled.Total control carriers are predicted to be more adaptable which renders active monitoring more promising.

The selection of the deploying location in an urban setting presents another challenge because getting local council approval may take some time. The contaminants have a stronger propensity to fall to the ground as a result of gravitational factors.When the sensor nodes are close to the efficient release height of pollution sources, it is thought that their installation cost is small. As a result, node placement is a crucial factor that must be taken into account. In massive scale uses, there may be hundreds of sensor nodes, and add-on capability is essential in such circumstances. It is observed that the majority of monitoring nodes lack this feature. Additionally required are flexible sensing and remote programming, sending intervals, and failure checking capability. A few additional factors, such as the lack of active surveillance, scalability, node utilisation of energy, 3D data acquisition, and mechanical and communication hardenings, also require attention[25].

8. AIR QUALITY PREDICTION METHODS

In the literature, a variety of techniques have been used, from statistical methods to more contemporary developments in machine learning. The most prevalent neural network topologies in the literature are deep feedforward and recurrent neural networks. Recurrent neural networks, which have a more specialised design, have proven to be an effective tool for time series projection. Deep learning has shown great effectiveness in discovering hidden relations within difficult issues. Ensemble learning is also advantageous because it is susceptible to noise and volatility.

8.1. Recurrent Neural Network

RNN variations are used in a variety of studies to capture temporal relationships. An LSTM model is used by Wang and Song (2018) to learn short-term as well as long-term temporal dependencies using the weather forecast[41]. K ok et al. (2017) use IoT sensor data with an LSTM solution to anticipate the near future[42]. When comparing the performance of several RNN cells, Athira et al. (2018) findsthat the GRU cell is marginally more accurate at determining PM10 concentration[43]. In order to estimate air quality concentrations, Li et al. (2017) used an LSTM model that takes spatio-temporal relationships into account. The extended LSTM (MAPE=11.93%) and SVR (MAPE=28.45%) findings show that the deep learning-based models perform better in terms of prediction[44].

8.2. Artificial Neural Network

According to clustering, Tamas et al. (2016) design different specialised MLP networks for every kind of weather. They also learn how to better categorise abrupt spikes by understanding the relationship among an elevated level of air pollutants and various weather classes[45].

In Ghoneim et al.'s (2017) study, the authors demonstrate that a deep learning regression technique can pick up patterns in the weather and pollution data gathered from 449 sensors spread throughout the Danish city of Aarhus. In terms of hourly prediction, their DNN model ($R2=0.91$) outperforms SVM ($R2=0.74$)[46].

ANN-inspired models perform exceptionally well because they have the ability to discover hidden relationships between spatial and temporal shape. Finding an ideal solution is made more challenging by the numerous drawbacks of ANN models, including the high dimension of hyperparameters. The over-fitting issue, which limits the generalisation of the ANN, is another drawback. Implementing well-known regularisation techniques, identifying the appropriate architecture for the problem, and optimising the hyperparameters for the learner's benefit are ways to get around these issues.

Journal of Chemical Health Risks

www.jchr.org

JCHR (2023) 13(5), 278-301 | ISSN:2251-6727

8.3. Ensemble Learning

The outcomes from an ensemble of an RF and FFNN were inferred using fuzzy logic by Bougoudis et al. (2016)[47]. In order to generalise the findings, they combine the strength of non-linear connection in a neural network with the averaging techniques of an ensemble approach. A RF model was also used by Chen et al. (2018) to forecast PM2.5 with features such as additional contaminants and weather variables. Compared to their generalised additive model implementation ($R2 = 0.55$), their RF model ($R2=0.83$) performs better[48].

Zhang et al. (2017) offer a solution that places more emphasis on the temporal variations in air pollution data samples. They address the issue of the intrinsic heterogeneity that various data samples across geography and time have[49]. The most appropriate model for each sample is determined by weighing an array of base-learners (RF, NN, KNN, SVM, and three knowledge-driven models) against it. In comparison to other techniques (such as stacking, adaBoost, bagging, and each base-learner), this huge multi-channel ensemble performs better.

9. MACHINE LEARNING (ML)

The goal of machine learning (ML) is to educate computers to automatically improve their ability to perform over time without being programmed specifically for a particular activity. The machine is taught using one or more of the available algorithms, trained on a specific quantity of information known as train information, and afterwards evaluated on a completely fresh set of information that it has never seen prior to, known as test information. Pattern recognition, picture analysis, forecasts, recommendation, and other uses for machine learning are some examples[50].

9.1. Supervised Learning

Labelled data are employed in supervised learning for the purpose of training. Regression and classification algorithms are components of supervised learning. The major purposes of regression algorithms are entity prediction and relational analysis of quantitative data. It explicitly uses Bayesian networks, fuzzy classification, decision trees, and linear regression. The set of data is divided into new classes using classification techniques, such as Support Vector Machines (SVM), Classification Tree, Random Forest, and Logistic Regression [50,51].

9.2. Unsupervised Learning

It is a direct contradiction to supervised learning. The training set does not contain any solved or labelled data. The machine must figure out the answer. It mostly consists of techniques for Dimension Reduction and Clustering. The core component is maintained while the enormous dataset with many features is reduced by dimension reduction algorithms into data with fewer features. Dimension Reduction includes Random Projection, Multidimensional Statistics, Tensor Reduction, and Principal Component Analysis (PCA). Contrarily, clustering algorithms divide the input data into groups based on specific criteria. Hierarchical clustering, Genetic algorithms, K-means clustering, and Gaussian mixture models are all included [50,51].

9.3. Reinforcement Learning

Using reinforcement learning, the agent realises to choose over time by the outcomes. It learns by trial-and-error and feedback techniques. Similar to Supervised Learning, it makes use of the input-output mapping but provides feedback through punishment and rewards. Markov Decision Process and Q Learning are the two crucial models for reinforcement learning. Self-driving cars, video games, recommendation systems, finance, trading, and other areas are examples of uses for reinforcement learning [52].

9.4. Machine Learning over Classical Statistical Forecasting Methods

Univariate time series issues are dealt with using conventional statistical forecasting techniques. Here are a few examples of traditional time series statistical techniques[52–54]:

i. **Naïve 2:** When the data contains lengthy periods of seeming ups and downs, this strategy, which resets the forecast to the most recent observation, is typically effective. Because of this, it is sometimes referred to as the optimised random walk model for the season.

ii. **Simple Exponential Smoothing (SES):** This approach is appropriate for datasets without obvious seasonality or trends.

Journal of Chemical Health Risks

www.jchr.org

JCHR (2023) 13(5), 278-301 | ISSN:2251-6727

iii. **Holt:** For data with trends, it is a supplemental measure of SES, and the Holt-Winter approach is employed to capture seasonality.

iv. **Damped Exponential Smoothing (DES):** Damped exponential smoothing is used to get rid of the concept that the trend will continue forever, as implied by earlier approaches.

v. **Theta Method:** The original data are divided up using this method into theta lines based on seasonality.

vi. **ARIMA:** The Auto-Regressive Integrated Moving Average, often known as ARIMA, seeks to link features. It is a method that is frequently used to solve time series issues.

vii. **ETS:** Error, Trend, and Seasonality is abbreviated as ET. Decomposition plots are used in an exponential smoothing model to determine whether to add, multiply, or ignore these trends, errors, and seasonality.

viii. **HMM:** The term "HMM" refers to a statistical Markov model that finds hidden states to understand Markov Chains. The probabilistic features of random processes are derived using this approach. Part of Speech Tagging is one of its key applications in the field of processing natural languages.

While machine learning algorithms-based techniques can detect the non-linear behaviour of the information set without knowing any other details about the dataset, classical models can only provide a linear behaviour among the target and the independent feature. These ETS or ARIMA models of data are also each time-series model's local model. ML-based models, on the other hand, combine learning throughout the entire series. As a result, when faced with a multivariate dataset, machine learning can be preferable to traditional methods. The move of the time series model from the traditional approach to these Machine Learning-based techniques has been driven by spectacular outcomes in other models, such as predicting shoreline evolution for sandy shores or forecasting exchange rates [55].

10. MACHINE LEARNING MODULE

The machine learning strategy is an iterative procedure that involves testing several solution possibilities. The implementation could simply be extended or replaced by fresh ideas, and fresh outcomes can be examined. The machine learning modules are in charge of tasks linked to machine learning. This module manages data preparation, designing features, hyperparameter search, training, assessment, and visualisation[56].

10.1. Data pre-processing

Data processing entails putting the raw information into a format that can be understood. Data imputation, which fills in missing values with substitute values, is the initial phase. In the datasets gathered, missing information from sensor readings are a common occurrence because to malfunction. The method of filling in the missing data is based on Lin et al. (2018)'s research and uses the average of nearby stations that are in close proximity. The mistake induced by dummy information is modest, and the surrounding stations' patterns resemble those of the one with the missing data. By calculating the weekly average of all neighbouring stations during the period of the missing measurement, the remaining missing values are filled in. To enable the feature engineering approach outlined below, all missing variables should be filled in rather to being ignored. To contain the temporal and statistical properties, the time series must be a continuous sequence. Although bias is incorporated into the models using the data imputation technique, this is thought to be less of an issue than just disregarding them. Finally, all pollutants data are trimmed for a minimum at zero to eliminate any potential negative sensor data values. The models' evaluation does not add any missing values. Instead, they disregard such timestamps' predictions. As a result, the evaluation in relation to actual observations will be greater[56].

10.2. Feature Engineering

A key component of machine learning that helps a learning algorithm function well is feature engineering. To create new, pertinent features that are presented in a way that the algorithms can understand, a thorough study of the unprocessed information is necessary. Creating robust and,

Journal of Chemical Health Risks
www.jchr.org
JCHR (2023) 13(5), 278-301 | ISSN:2251-6727

preferably, straightforward links among newly acquired input features and the output target is the aim of feature engineering. In order to make sense of the data's complexity, underlying domain knowledge is used. The date and time of the measurement is mostly used to construct the temporal characteristics. The date and time contain details on the day, week, month, day of the month, and seasonality of the year. To determine whether a given day is a holiday in country, the date is compared to the calendar. The final temporal feature is produced using the parameter's previous values.

The geographic features include characteristics from nearby stations. These are determined using the average of the closest pollution stations. To aid the models in capturing the spatial relationships of the air quality, several elements have been added. By using a collection of mathematical operations on the time period to deduce special qualities, statistical characteristics are created. Instead of only including historical values, statistical characteristics aim to offer a more generic and extensive temporal dependency. The statistical functions will include a more intelligent relationship from the past that the models can learn more quickly.

The statistical aspects will offer trustworthy and simpler connections between the past and the predictions. In order to understand the complexity, it can be helpful to smooth the time series' raw values using statistical feature engineering. The lowest, highest, and moving average functions can primarily aid in capturing patterns within the series. By discovering what occurred soon before the change, the variation and divergence can be used to detect rapid changes. Because fresh data points are less likely to be identical to previously learned features, the huge feature database may make it difficult for machine learning approaches to learn. It becomes more difficult for the algorithm to generalise well when there is a likelihood of fewer connections among the past and the future. The likelihood of overfitting to noise increases along with the variation, which lowers performance. Multiple regularisation techniques, however, are used to circumvent this restriction of a greater variety of features in order to take overfitting into consideration[56].

10.3. Dataset Split

The dataset is divided into three sections for the purpose of assessing the performance of the model. 80% of the authentic collection is chosen for training, a small portion of it is selected for approval, and the rest of it is chosen for testing. The time series is kept continuous for this initial split, and the most recent measurements are used for the testing portion. The entire dataset is compiled over many years. The testing data will then consist of slightly more than a year's worth of observations, whereas the training data will consist about of hourly data points. When the machine learning algorithms are trained on the entire dataset before being deployed in production, the option to evaluate the models in this way is made to gain a solid indication of the way the algorithms will operate in the future. There haven't been any notable alterations to the distribution across the given time period of air quality observations. So long as the distribution stays steady, the information can be used to make predictions without risk. However, since projections are intended to prevent air pollution peaks, they will be affected if the data on air quality significantly change as a result of any city-taken interventions. Additionally, the estimates derived from the test results can be contrasted with official national projections for air quality. The ten percent of the set of training data is provided for validation while the remaining 90% is used to train all the models. The validation set aids in preventing overfitting throughout the training process[56].

10.4. Feature Scaling

No scaling is done for the gradient boosting, random forest, or ridge approaches because it barely affects the outcome. By supporting the network's activation processes, the features in neural networks are scaled to shorten training times. The activation functions ReLU and LeakyReLU used by neural networks perform best when the allowed values are above 0 to prevent disappearing gradients and below 1 to prevent exploding gradients[56].

10.5. Hyperparemeter Search

For each model, a number of hyperparameters were optimised using various optimisation techniques. A random search was used to find the hyperparameters. The wide range

Journal of Chemical Health Risks
www.jchr.org
JCHR (2023) 13(5), 278-301 | ISSN:2251-6727

of parameters of the hyperparameters was reduced by the random parameter search.

Bergstra and Bengio (2012) demonstrated that an arbitrary hyperparameter search is both conceptually and experimentally more effective than a grid search. To train the model, the best collection of hyperparameters are selected. During the search, a high correlation among the model and the window size's hyperparameters was discovered. Thus, a window size of 24 hours is selected for the hyperparameter search. The outcomes of the hyperparameter tuning for neural networks for each type of pollutant changed sufficiently for three separate sets of hyperparameters to be created for each pollutant (PM2.5, PM10, and NO2)[56].

11. BIG DATA

Big data is not just data that is large in size and needs a lot of storage; in fact, this is one of its key characteristics. The volume of this complicated, unstructured, and quickly producing data is enormous. Big data has been defined in a few different ways. The V's concepts are typically used to define big data. Originally established by Doug Laney with the three V's notions of volume, velocity, and variety, it was then expanded to add the fourth V: veracity. Currently, it has more than twenty Vs. The scalability and effectiveness of the conventional ML and data analysis techniques for big data were among the new problems imposed by big data. The answer is provided by big data analytics, which uses these conventional techniques on diverse big data platforms[57]. The concept of big data is shown in Figure 7

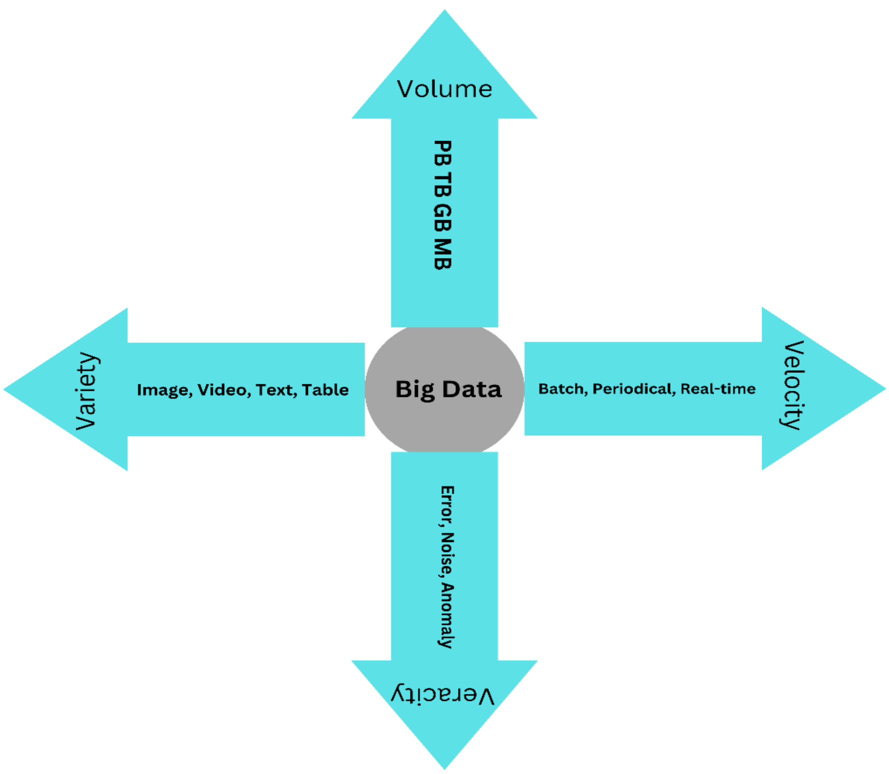

Figure 7 Perception of Big Data

11.1. Big Data and Machine Learning-Based Monitoring and Forecasting Systems

To produce more accurate and dependable results, the data downloaded or retrieved through a real-time method must go through pre-processing. Noise, missing values or attributes, or errors can all be found in the data that was gathered from sources. Data cleaning and data pre-processing are terms used to describe the process of cleansing the raw data. Big data technologies are used to manage this enormous amount of data after it has been cleansed and integrated from all of the sources. Utilising machine learning methods, target data mining is carried out once all high-quality data has been extracted.

Borges et al. used four IoT installations to gather the CO data for the Sao Paulo Metropolitan Area (SMA), and R studio was used to send the data to Apache Hadoop. To handle and analyse the data, MapReduce was used, which is a popular HDFS (Hadoop Distributed File System) technology. Shiny is a R tool that creates interactive web applications by visualising data. Data from all the sensor configurations were compared, and Shiny was used to plot the sensor density, mean values for each sensor each day, and a summary of the measurements. The CO value for the SMA was discovered to be twice what the WHO had set[58].

Real-time monitoring and prediction system was suggested by Ayele et al. The Long Short-Term Memory (LSTM) method was used to make predictions after data from the IoT setup was collected and stored on the web server. With TensorFlow serving as the backend, this experiment was carried out in Python 3.6.3. An accurate conclusion to this experiment was reached[59].

By using an IoT setup to gather data, Moses et al. provided users with a Google Map suggestion for a different, less polluting route. The neural network (NN) and support vector machine (SVM) regression algorithm were used to forecast the air quality over time. SVM promotes error tolerance by individualising hyperplanes, whereas NN with sigmoid activation function was utilised for AQI value prediction[60].

The Central Pollution Control Board (CPCB) threshold values were taken into account by Srivastava et al. when using IoT setup to collect data and transmit emergency notifications via an Android application. Two techniques, Random Forest Regression (RFR) and Support Vector Regression (SVR), perform predictive analysis. Precision, recall, support, and F1 score were used in the models that correspond to these algorithms. In the RFR model, accuracy was close to 99%, compared to 90% in the SVR model. Because the algorithm makes use of numerous different decision trees, the performance of RFR was good[61].

A system for data evaluation and prediction of air pollution was suggested by Wang et al.. Using an API key, the data was obtained from the Chinese air quality online monitoring and analysis portal. HDFS was utilised to store the atmospheric data, and Spark was employed as the calculating engine for the industrial data. To get rid of redundant or incorrect data, preprocessing was done on the data. The Backward Propagation (BP) neural network was used for data mining. A supervised learning method known as BP is frequently used for forecasting so that output is much more closely related to the predicted vector and weights can be modified regularly. Data from the BP neural network, including AQI, CO, NO2, PM2.5, and PM10 as well as city, date, and wind speed, were gathered using the Python crawler. It was discovered that the wind speed measurement is one of the crucial components. A visualisation tool was used to present the outcomes. This system proposed an effective way by combining a number of technologies[62].

A effective deep learning approach for smart cities has been presented by Kök et al. The CityPulse EU FP7 Project of the cities of Aarhus and Brasov in Denmark and Romania, respectively, provided the data that was utilised. For this research, ozone and NO2 pollution were taken into account. Python platforms Keras DL framework and TensorFlow were employed, and the dataset was split into two portions of 69.5% and 30%, respectively. SVM and LSTM, two prediction algorithms, were trained on the data. A well-known Recurrent Neural Network (RNN) with feedback connections is the LSTM. Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) are used to compare the two models, along with each model's Confusion Matrix, Precision, Recall, and F1 Score. Yellow and green

Journal of Chemical Health Risks

www.jchr.org

JCHR (2023) 13(5), 278-301 | ISSN:2251-6727

are associated with the model that makes almost accurate forecasts for red alarms. The best precision and F1-Score were obtained by LSTM and were 98% and 97%, respectively, whereas for SVM, the values for red alarm were 95% (precision) and 96% (F1 score). For IoT data gathered through the CityPulse EU FP7 Project, this solution produced effective and encouraging results[63].

The Central Pollution Control Boards (CPCB) and the Meteorological department provided Nandini and Fathima with pollution and meteorological data. The dataset was then split into test data and validation data, with 90% and 10%, respectively. The data was then divided into three clusters using K-Means Clustering: Good, Moderate, and Unhealthy. The relationship between factors and results was then observed using multinominal logistic regression, and conditional control statements were improved using a decision tree as a support tool. The effectiveness of both algorithms is evaluated using a confusion matrix. The decision tree model's error rate was found to be 0.666, while the regression model's was 0.428. As a result, the regression model provided the best fit[64].

Mahalingam et al. also investigated the effectiveness of machine learning techniques. Dataset for New Delhi, India, was collected from the Central Pollution Control Boards (CPCB). Following that, the entire set of data was classified into six categories: good, satisfactory, moderate, poor, very poor, and severe. With eight neurons in the hidden layer and six different types of Support Vector Machines (SVM), neural networks perform prediction analysis. Linear SVM (89.2%), Quadratic SVM (94.6%), Cubic SVM (94.6%), Fine Gaussian (62.2%), Medium Gaussian (97.3%), and Coarse Gaussian (78.4%) are the six SVMs with the highest accuracy. The accuracy of neural networks is 91.62%. Medium Gaussian is therefore determined to be the best fit[65].

12. RESEARCH ISSUES AND CHALLENGES

Some research difficulties and obstacles for creating and putting into practise effective models for air pollution surveillance and forecasting are highlighted after analysing and analysing the literature in the aforementioned sections. Whether intentionally or unintentionally, forecasting is a technique used to make predictions about the future and the possibility of particular events. Future research may go in the directions covered by the issues and difficulties with the research that are addressed below.

(i) Quality of Data: The IoT infrastructure gathers the data, however occasionally the data quality deteriorates because of poor network, sensor, and connection quality. Results will be more accurately produced if the data quality is high and there are fewer missing and incorrect items.

(ii) Quantity of Data: For the model's correctness, the information contained in records and quantity taken into account are also crucial. Testing the models on various data sets is necessary. For a better model, the right quantity of information must be taken into account.

(iii) Real-Time Integrated Model: Currently, there is still a dearth of an integrated big data and machine learning-based system for monitoring and predicting air quality in real-time. The amount of pollutants in the air can vary depending on where you are. Unstable and integrated modelling are still required to address all the dynamic changes.

(iv) Meteorological Factors: By combining meteorological data for that region with pollution data, one can increase the significance of meteorological components in models and the accuracy of those models.

(v) Uniformity of Sensor Setups: The quality of the data is impacted by the non-uniform deployment of sensors throughout the cities. To monitor and analyse the city's air quality, it is critical that sensor installations be consistent throughout.

(vi) Number of Sensors Setups: A significant number of IoT installations must be put around the area in order to collect data on air quality. This will increase both the model's general stability and the data's suitability for mining.

(vii) Processing Time of Models: Another crucial element is how quickly machine learning models are processed. Machine learning algorithms were only assessed on their accuracies and mistakes in a select few proposed systems. An effective air pollution prediction and tracking model should use an algorithm that operates with excellent precision and quick processing times.

(viii) Number of Pollutants Taken into Account: The concentration of different hazardous gases and particles affects the region's air quality. Some studies used a small number of gases as an input when calculating AQI. In a perfect world, every gas that contributes to air pollution would be considered. To account for all of the airborne toxins, an effective model is still needed.

(ix) Checking Correlations: If the association among features can be found before modelling it, machine learning algorithms can be enhanced. The effectiveness of the model can be enhanced by removing the features that have a weak link with the target feature.

(x) Hyperparameter Tuning: Every piece of data exhibits a unique behaviour, therefore tweaking the hyperparameters may aid in improving the ML model's efficiency[57].

13. CONCLUSION AND FUTURE WORK

The effects of overusing nature are returning to earth's inhabitants in the way of lethal consequences. Air quality requires ongoing monitoring and assessment. The Internet of Things, Big Data methods, and machine learning are examples of modern interaction, computation, and analytics technologies that could give us more reliable and effective approaches for tracking and predicting air pollution. The contemporary approach to air pollution surveillance and forecasting systems has been discussed and examined in this research. Additionally, the model and systems now in use are compared on a number of factors, and this study discusses some important research challenges. Additionally, several doable strategies for enhancing the models are proposed. Currently, the range, precision, and efficiency of air pollution detecting and monitoring methods and procedures are limited. Future work will involve setting up of the models that are available and real-time monitoring. The models based on data can be created to forecast, suggest, and track future work in order to control sicknesses and climate change. Finally, reducing air pollution can be greatly aided by prevention and education. Although they won't stop the negative consequences, education and prevention efforts will make things sustainable going forward.

Acknowledgements:

Authors acknowledge the support of staff of ISBM University, Gariyaband, Chhattisgarh for their support, valuable guidance and encouragement.

Funding:

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Disclosure statement

The authors report there are no competing interests to declare.

REFERENCES

[1] R. Turnbull, K. Rogers, A. Martin, M. Rattenbury, R. Morgan, Human impacts recorded in chemical and isotopic fingerprints of soils from Dunedin City, New Zealand, Science of The Total Environment. 673 (2019) 455–469. https://doi.org/10.1016/J.SCITOTENV.2019.04.063.

[2] S. Fahad, O. Sönmez, S. Saud, D. Wang, C. Wu, M. Adnan, V. Turan, Plant Growth Regulators for Climate-Smart Agriculture, Plant Growth Regulators for Climate-Smart Agriculture. (2021). https://doi.org/10.1201/9781003109013/PLANT-GROWTH-REGULATORS-CLIMATE-SMART-AGRICULTURE-SHAH-FAHAD-SHAH-SAUD-CHAO-WU-DEPENG-WANG-OSMAN-SONMEZ-MUHAMMAD-ADNAN-VEYSEL-TURAN.

[3] O. Sönmez, S. Saud, D. Wang, C. Wu, M. Adnan, V. Turan, Climate Change and Plants, (2021). https://doi.org/10.1201/9781003108931.

[4] Y. Rybarczyk, R. Zalakeviciute, Assessing the COVID-19 Impact on Air Quality: A Machine Learning Approach, Geophys Res Lett. 48 (2021) e2020GL091202. https://doi.org/10.1029/2020GL091202.

[5] S. Gulia, S.M. Shiva Nagendra, M. Khare, I. Khanna, Urban air quality management-A review, Atmos Pollut Res. 6 (2015) 286–304. https://doi.org/10.5094/APR.2015.033.

[6] R. Agarwal, G. Jayaraman, S. Anand, P. Marimuthu, Assessing respiratory morbidity through pollution

status and meteorological conditions for Delhi, Environ Monit Assess. 114 (2006) 489–504. https://doi.org/10.1007/S10661-006-4935-3/METRICS.

[7] W.Y. Yi, K.M. Lo, T. Mak, K.S. Leung, Y. Leung, M.L. Meng, A Survey of Wireless Sensor Network Based Air Pollution Monitoring Systems, Sensors 2015, Vol. 15, Pages 31392-31427. 15 (2015) 31392–31427. https://doi.org/10.3390/S151229859.

[8] M. Chiang, T. Zhang, Fog and IoT: An Overview of Research Opportunities, IEEE Internet Things J. 3 (2016) 854–864. https://doi.org/10.1109/JIOT.2016.2584538.

[9] M.F.M. Firdhous, B.H. Sudantha, P.M. Karunaratne, IoT enabled proactive indoor air quality monitoring system for sustainable health management, Proceedings of the 2017 2nd International Conference on Computing and Communications Technologies, ICCCT 2017. (2017) 216–221. https://doi.org/10.1109/ICCCT2.2017.7972281.

[10] L. Moses, Tamilselvan, Raju, Karthikeyan, IoT enabled Environmental Air Pollution Monitoring and Rerouting system using Machine learning algorithms, IOP Conf Ser Mater Sci Eng. 955 (2020) 012005. https://doi.org/10.1088/1757-899X/955/1/012005.

[11] R. Khot, V. Chitre, Survey on air pollution monitoring systems, Proceedings of 2017 International Conference on Innovations in Information, Embedded and Communication Systems, ICIIECS 2017. 2018-January (2018) 1–4. https://doi.org/10.1109/ICIIECS.2017.8275846.

[12] H. Akimoto, Global Air Quality and Pollution, Science (1979). 302 (2003) 1716–1719. https://doi.org/10.1126/SCIENCE.1092666/SUPPL_FILE/COVER.GIF.

[13] Y.C. Liang, Y. Maimury, A.H.L. Chen, J.R.C. Juarez, Machine Learning-Based Prediction of Air Quality, Applied Sciences 2020, Vol. 10, Page 9151. 10 (2020) 9151. https://doi.org/10.3390/APP10249151.

[14] S. Mansour, N. Nasser, L. Karim, A. Ali, Wireless sensor network-based air quality monitoring system, 2014 International Conference on Computing, Networking and Communications, ICNC 2014. (2014) 545–550. https://doi.org/10.1109/ICCNC.2014.6785394.

[15] K.T. Min, A. Forys, T. Schmid, Demonstration abstract: AirFeed - Indoor real time interactive air quality monitoring system, IPSN 2014 - Proceedings of the 13th International Symposium on Information Processing in Sensor Networks (Part of CPS Week). (2014) 325–326. https://doi.org/10.1109/IPSN.2014.6846785.

[16] L. Spinelle, M. Gerboles, G. Kok, S. Persijn, T. Sauerwald, Review of Portable and Low-Cost Sensors for the Ambient Air Monitoring of Benzene and Other Volatile Organic Compounds, Sensors 2017, Vol. 17, Page 1520. 17 (2017) 1520. https://doi.org/10.3390/S17071520.

[17] J.Y. Kim, C.H. Chu, S.M. Shin, ISSAQ: An integrated sensing systems for real-time indoor air quality monitoring, IEEE Sens J. 14 (2014) 4230–4244. https://doi.org/10.1109/JSEN.2014.2359832.

[18] J. Kang, K. Il Hwang, A Comprehensive Real-Time Indoor Air-Quality Level Indicator, Sustainability 2016, Vol. 8, Page 881. 8 (2016) 881. https://doi.org/10.3390/SU8090881.

[19] A. Boubrima, W. Bechkit, H. Rivano, Optimal WSN Deployment Models for Air Pollution Monitoring, IEEE Trans Wirel Commun. 16 (2017) 2723–2735. https://doi.org/10.1109/TWC.2017.2658601.

[20] A.C. Rai, P. Kumar, F. Pilla, A.N. Skouloudis, S. Di Sabatino, C. Ratti, A. Yasar, D. Rickerby, End-user perspective of low-cost sensors for outdoor air pollution monitoring, Science of The Total Environment. 607–608 (2017) 691–705. https://doi.org/10.1016/J.SCITOTENV.2017.06.266.

[21] Z. Idrees, Z. Zou, L. Zheng, Edge Computing Based IoT Architecture for Low Cost Air Pollution Monitoring Systems: A Comprehensive System Analysis, Design Considerations & Development, Sensors 2018, Vol. 18, Page 3021. 18 (2018) 3021. https://doi.org/10.3390/S18093021.

[22] K.B. Swain, G. Santamanyu, A.R. Senapati, Smart industry pollution monitoring and controlling using LabVIEW based IoT, Proceedings of 2017 3rd IEEE

International Conference on Sensing, Signal Processing and Security, ICSSS 2017. (2017) 74–78. https://doi.org/10.1109/SSPS.2017.8071568.

[23] X. Yang, L. Yang, J. Zhang, A WiFi-enabled indoor air quality monitoring and control system: The design and control experiments, IEEE International Conference on Control and Automation, ICCA. (2017) 927–932. https://doi.org/10.1109/ICCA.2017.8003185.

[24] P. Bellavista, C. Giannelli, R. Zamagna, The PeRvasive Environment Sensing and Sharing Solution, Sustainability 2017, Vol. 9, Page 585. 9 (2017) 585. https://doi.org/10.3390/SU9040585.

[25] Z. Idrees, L. Zheng, Low cost air pollution monitoring systems: A review of protocols and enabling technologies, J Ind Inf Integr. 17 (2020) 100123. https://doi.org/10.1016/J.JII.2019.100123.

[26] R. Baron, J. Saffell, Amperometric Gas Sensors as a Low Cost Emerging Technology Platform for Air Quality Monitoring Applications: A Review, ACS Sens. 2 (2017) 1553–1566. https://doi.org/10.1021/ACSSENSORS.7B00620/ASSET/IMAGES/MEDIUM/SE-2017-00620N_0018.GIF.

[27] H. Li, Y. Zhu, Y. Zhao, T. Chen, Y. Jiang, Y. Shan, Y. Liu, J. Mu, X. Yin, D. Wu, C. Zhang, S. Si, X. Wang, W. Wang, L. Xue, Evaluation of the Performance of Low-Cost Air Quality Sensors at a High Mountain Station with Complex Meteorological Conditions, Atmosphere 2020, Vol. 11, Page 212. 11 (2020) 212. https://doi.org/10.3390/ATMOS11020212.

[28] B. Maag, Z. Zhou, L. Thiele, A Survey on Sensor Calibration in Air Pollution Monitoring Deployments, IEEE Internet Things J. 5 (2018) 4857–4870. https://doi.org/10.1109/JIOT.2018.2853660.

[29] D. Hasenfratz, O. Saukh, C. Walser, C. Hueglin, M. Fierz, T. Arn, J. Beutel, L. Thiele, Deriving high-resolution urban air pollution maps using mobile sensor nodes, Pervasive Mob Comput. 16 (2015) 268–285. https://doi.org/10.1016/J.PMCJ.2014.11.008.

[30] A. Arfire, A. Marjovi, A. Martinoli, Model-based rendezvous calibration of mobile sensor networks for monitoring air quality, 2015 IEEE SENSORS - Proceedings. (2015). https://doi.org/10.1109/ICSENS.2015.7370258.

[31] H. Hojaiji, O. Goldstein, C.E. King, M. Sarrafzadeh, M. Jerrett, Design and calibration of a wearable and wireless research grade air quality monitoring system for real-time data collection, GHTC 2017 - IEEE Global Humanitarian Technology Conference, Proceedings. 2017-January (2017) 1–10. https://doi.org/10.1109/GHTC.2017.8239308.

[32] H.I. Kobo, A.M. Abu-Mahfouz, G.P. Hancke, A Survey on Software-Defined Wireless Sensor Networks: Challenges and Design Requirements, IEEE Access. 5 (2017) 1872–1899. https://doi.org/10.1109/ACCESS.2017.2666200.

[33] S. Lee, J. Jo, Y. Kim, H. Stephen, A framework for environmental monitoring with arduino-based sensors using restful web service, Proceedings - 2014 IEEE International Conference on Services Computing, SCC 2014. (2014) 275–282. https://doi.org/10.1109/SCC.2014.44.

[34] X. Fang, I. Bate, Issues of using wireless sensor network to monitor urban air quality, FAILSAFE 2017 - Proceedings of the 1st ACM International Workshop on the Engineering of Reliable, Robust, and Secure Embedded Wireless Sensing Systems, Part of SenSys 2017. (2017) 32–39. https://doi.org/10.1145/3143337.3143339.

[35] C. Xiaojun, L. Xianpeng, X. Peng, IOT-based air pollution monitoring and forecasting system, 2015 International Conference on Computer and Computational Sciences, ICCCS 2015. (2015) 257–260. https://doi.org/10.1109/ICCACS.2015.7361361.

[36] S.H. Kim, J.M. Jeong, M.T. Hwang, C.S. Kang, Development of an IoT-based atmospheric environment monitoring system, International Conference on Information and Communication Technology Convergence: ICT Convergence Technologies Leading the Fourth Industrial Revolution, ICTC 2017. 2017-December (2017)

861–863. https://doi.org/10.1109/ICTC.2017.8190799.

[37] M. Yokoyama, T. Hara, S.K. Madria, Efficient diversified set monitoring for mobile sensor stream environments, Proceedings - 2017 IEEE International Conference on Big Data, Big Data 2017. 2018-January (2017) 500–507. https://doi.org/10.1109/BIGDATA.2017.8257964.

[38] S.C. Hu, Y.C. Wang, C.Y. Huang, Y.C. Tseng, Measuring air quality in city areas by vehicular wireless sensor networks, Journal of Systems and Software. 84 (2011) 2005–2012. https://doi.org/10.1016/J.JSS.2011.06.043.

[39] W. Jiao, G. Hagler, R. Williams, R. Sharpe, R. Brown, D. Garver, R. Judge, M. Caudill, J. Rickard, M. Davis, L. Weinstock, S. Zimmer-Dauphinee, K. Buckley, Community Air Sensor Network (CAIRSENSE) project: Evaluation of low-cost sensor performance in a suburban environment in the southeastern United States, Atmos Meas Tech. 9 (2016) 5281–5292. https://doi.org/10.5194/AMT-9-5281-2016.

[40] M.T. Lazarescu, Design of a WSN platform for long-term environmental monitoring for IoT applications, IEEE J Emerg Sel Top Circuits Syst. 3 (2013) 45–54. https://doi.org/10.1109/JETCAS.2013.2243032.

[41] J. Wang, G. Song, A Deep Spatial-Temporal Ensemble Model for Air Quality Prediction, Neurocomputing. 314 (2018) 198–206. https://doi.org/10.1016/J.NEUCOM.2018.06.049.

[42] I. Kök, M.U. Şimşek, S. Özdemir, A deep learning model for air quality prediction in smart cities, Proceedings - 2017 IEEE International Conference on Big Data, Big Data 2017. 2018-January (2017) 1983–1990. https://doi.org/10.1109/BIGDATA.2017.8258144.

[43] V. Athira, P. Geetha, R. Vinayakumar, K.P. Soman, DeepAirNet: Applying Recurrent Networks for Air Quality Prediction, Procedia Comput Sci. 132 (2018) 1394–1403. https://doi.org/10.1016/J.PROCS.2018.05.068.

[44] X. Li, L. Peng, X. Yao, S. Cui, Y. Hu, C. You, T. Chi, Long short-term memory neural network for air pollutant concentration predictions: Method development and evaluation, Environmental Pollution. 231 (2017) 997–1004. https://doi.org/10.1016/J.ENVPOL.2017.08.114.

[45] W. Tamas, G. Notton, C. Paoli, M.L. Nivet, C. Voyant, Hybridization of Air Quality Forecasting Models Using Machine Learning and Clustering: An Original Approach to Detect Pollutant Peaks, Aerosol Air Qual Res. 16 (2016) 405–416. https://doi.org/10.4209/AAQR.2015.03.0193.

[46] O.A. Ghoneim, Doreswamy, B.R. Manjunatha, Forecasting of ozone concentration in smart city using deep learning, 2017 International Conference on Advances in Computing, Communications and Informatics, ICACCI 2017. 2017-January (2017) 1320–1326. https://doi.org/10.1109/ICACCI.2017.8126024.

[47] I. Bougoudis, K. Demertzis, L. Iliadis, HISYCOL a hybrid computational intelligence system for combined machine learning: the case of air pollution modeling in Athens, Neural Comput Appl. 27 (2015) 1191–1206. https://www.academia.edu/81017779/HISYCOL_a_hybrid_computational_intelligence_system_for_combined_machine_learning_the_case_of_air_pollution_modeling_in_Athens (accessed May 26, 2023).

[48] G. Chen, S. Li, L.D. Knibbs, N.A.S. Hamm, W. Cao, T. Li, J. Guo, H. Ren, M.J. Abramson, Y. Guo, A machine learning method to estimate PM2.5 concentrations across China with remote sensing, meteorological and land use information, Science of The Total Environment. 636 (2018) 52–60. https://doi.org/10.1016/J.SCITOTENV.2018.04.251.

[49] C. Zhang, J. Yan, Y. Li, F. Sun, J. Yan, D. Zhang, X. Rui, R. Bie, Early Air Pollution Forecasting as a Service: An Ensemble Learning Approach, Proceedings - 2017 IEEE 24th International Conference on Web Services, ICWS 2017. (2017) 636–643. https://doi.org/10.1109/ICWS.2017.76.

[50] P. Louridas, C. Ebert, Machine Learning, IEEE Softw. 33 (2016) 110–115. https://doi.org/10.1109/MS.2016.114.

[51] M.R.M. Talabis, R. McPherson, I. Miyamoto, J.L. Martin, D. Kaye, Information Security Analytics: Finding Security Insights, Patterns and Anomalies in Big Data, Elsevier, 2014. https://doi.org/10.1016/B978-0-12-800207-0.01001-7.

[52] R.J. Hyndman, G. Athanasopoulos, Forecasting: principles and practice, 2018. https://books.google.co.in/books?hl=en&lr=&id=_bBhDwAAQBAJ&oi=fnd&pg=PA7&dq=R.+J.+Hyndman+and+G.+(2021)+%26+Athanasopoulos,+%E2%80%9CForecasting:+Principles+and+Practice+(3rd+ed).%E2%80%9D&ots=Tim_tj_NLO&sig=GoU5xGU-rz2y7TbiI9H1nODrA8U (accessed May 26, 2023).

[53] L.R. Rabiner, B.H. Juang, An Introduction to Hidden Markov Models, IEEE ASSP Magazine. 3 (1986) 4–16. https://doi.org/10.1109/MASSP.1986.1165342.

[54] Á. Gómez-Losada, J.C.M. Pires, R. Pino-Mejías, Modelling background air pollution exposure in urban environments: Implications for epidemiological research, Environmental Modelling & Software. 106 (2018) 13–21. https://doi.org/10.1016/J.ENVSOFT.2018.02.011.

[55] P.A. Sánchez-Sánchez, J.R. García-González, L.H.P. Coronell, P.A. Sánchez-Sánchez, J.R. García-González, L.H.P. Coronell, Encountered Problems of Time Series with Neural Networks: Models and Architectures, Recent Trends in Artificial Neural Networks - from Training to Prediction. (2019). https://doi.org/10.5772/INTECHOPEN.88901.

[56] A.J. Lepperød, A. Jacobsen, L. Ntnu, Air Quality Prediction with Machine Learning, (2019). https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/2626165 (accessed May 26, 2023).

[57] A. Gangwar, S. Singh, R. Mishra, S. Prakash, The State-of-the-Art in Air Pollution Monitoring and Forecasting Systems Using IoT, Big Data, and Machine Learning, Wirel Pers Commun. 130 (2023) 1699–1729. https://doi.org/10.1007/S11277-023-10351-1/METRICS.

[58] M.A. Borges, G.F. Melo, C. De Massaki, O. Igarashi, P.B. Lopes, L.A. Silva, An architecture for the internet of things and the use of big data techniques in the analysis of carbon monoxide, Proceedings - 2017 IEEE International Conference on Information Reuse and Integration, IRI 2017. 2017-January (2017) 184–191. https://doi.org/10.1109/IRI.2017.76.

[59] T.W. Ayele, R. Mehta, Air pollution monitoring and prediction using IoT, Proceedings of the International Conference on Inventive Communication and Computational Technologies, ICICCT 2018. (2018) 1741–1745. https://doi.org/10.1109/ICICCT.2018.8473272.

[60] J. Strake, F. Kaiser, F. Basiri, al -, D. Witthaut -, Y. Wang, Y. Liu, J. Huang, L. Moses, IoT enabled Environmental Air Pollution Monitoring and Rerouting system using Machine learning algorithms, IOP Conf Ser Mater Sci Eng. 955 (2020) 012005. https://doi.org/10.1088/1757-899X/955/1/012005.

[61] H. Srivastava, S. Mishra, S.K. Das, S. Sarkar, An IoT-Based Pollution Monitoring System Using Data Analytics Approach, Lecture Notes in Electrical Engineering. 686 (2020) 187–198. https://doi.org/10.1007/978-981-15-7031-5_18/COVER.

[62] J. Wang, L. Ai, Y. Wan, H. Chen, J. Zhang, X.G. Yuan, Y. Tang, H. Hao, Y. Zuo, Z. Tan, M. Qiao, Y.H. Cao, Air quality data analysis and forecasting platform based on big data, Proceedings - 2019 Chinese Automation Congress, CAC 2019. (2019) 2042–2046. https://doi.org/10.1109/CAC48633.2019.8996332.

[63] I. Kök, M.U. Şimşek, S. Özdemir, A deep learning model for air quality prediction in smart cities, Proceedings - 2017 IEEE International Conference on Big Data, Big Data 2017. 2018-January (2017) 1983–1990. https://doi.org/10.1109/BIGDATA.2017.8258144.

[64] K. Nandini, G. Fathima, Urban Air Quality Analysis and Prediction Using Machine Learning, 1st International Conference on Advanced Technologies

in Intelligent Control, Environment, Computing and Communication Engineering, ICATIECE 2019. (2019) 98–102. https://doi.org/10.1109/ICATIECE45860.2019.9063845.

[65] U. Mahalingam, K. Elangovan, H. Dobhal, C. Valliappa, S. Shrestha, G. Kedam, A machine learning model for air quality prediction for smart cities, 2019 International Conference on Wireless Communications, Signal Processing and Networking, WiSPNET 2019. (2019) 452–457. https://doi.org/10.1109/WISPNET45539.2019.9032734.