iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: VIII Month of publication: August 2025

DOI: https://doi.org/10.22214/ijraset.2025.73888

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue VIII Aug 2025- Available at www.ijraset.com

A Survey on Air Pollution Prediction Using Machine Learning Techniques

Mrs.P.Umasanthiya[1], Dr. Marrynal S. Eastaff[2]
[1]Assistant Professor, [2]Associate Professor

***Abstract:** Predicting air pollution levels is crucial because of the damage it may do to ecosystems and people's health. A comprehensive review of current methods and models for forecasting air pollution has been presented in this article. A thorough review of 32 scholarly publications explores various machine learning algorithms, statistical models, and hybrid approaches to predict future pollution concentrations accurately. We discuss the unique challenges of air quality prediction, including data variability, spatial-temporal correlations, and the effect of meteorological factors. The research divides existing prediction models into categories based on their methodologies, data requirements, and application scenarios, stressing their strengths and limitations. The study also looks at how new technologies like IoT sensors, deep learning, and ensemble techniques might improve the precision and dependability of air pollution predictions. Future research focuses on integrating real-time data, multi-source information fusion, and developing scalable, interpretable models for dynamic air quality management.*
***Keywords:** Air Pollution, Prediction, Machine Learning, Statistical Models, Forecasting.*

I. INTRODUCTION

Factors that contribute to air pollution include the number of people living in an area, the kind of industry they work in, the number of thermal power plants in the region, the energy sector, the automotive industry, and modes of mobility. Premature mortality, skin rashes, respiratory tract infections, pneumonia, lung cancer, and heart failure are only some of the human and environmental health problems caused by air pollution[1-4]. Worldwide, non-governmental organisations (NGOs) need to monitor air quality indicators (AQIs) for several important factors that define air pollution in a given area, such as particulate matter, gaseous pollutants, and metrological features [5-7].Several important factors that contribute to the air quality index include particulate matter 2.5 (PM2.5), particulate matter 10 (PM10), carbon dioxide (CO2), carbon monoxide (CO), sulfur oxides (SOx), nitrogen oxides (NOx), ozone (O3), and ammonia (NH3).Acid rain, increased AQI due to these compounds, smog and aerosol production, reduced visibility, and climate change are only a few ways these pollutants damage the ecosystem [8–10]. The main culprits in the rise in global temperatures are greenhouse gases (GHGs), impact plant-soil interactions and climate change, which has significant negative consequences for agriculture, the environment, and the economy[11-15]. Using data from 2010 to 2019, the World Health Organization (WHO) released a global air quality assessment in 2022. This research found that PM2.5 was growing worldwide and was responsible for 1.7 million annual deaths in India alone. It was based on investigations of 6743 cities in 117 different countries and assessed a broad spectrum of air pollutants indicated above.In the following years, India's air pollution will seriously affect public health, as eighteen of the country's cities are among the most polluted in the world. The most dangerous conditions for human health and safety are those with a high AQI. As a result, AQI forecasting and monitoring have emerged as critical tools for achieving global sustainability. Statistical, deterministic, physical, machine learning and deep learning models for AQI prediction have been developed by several authors [16–19]. The rigidity of complex situations makes them unsuitable for statistical and decision-making models. Thanks to recent sensor developments, it is now easy to detect different levels of air pollution and automatically compute AQI. It is easy to forecast AQI using freely accessible data sets. The machine learning system reliably and correctly predicts the AQI in all environmental conditions. Machine learning enables us to deliver more precise AQI forecasts as historical data becomes more readily available for analysis.
Their increasing popularity results from their attempts to displace more traditional statistical models for time-series prediction[20–24]. It is challenging to create a statistical model that can forecast such occurrences due to the poorly known dynamics of pollution concentrations and the highly nonlinear processes affecting them.By using historical data to discover the connection between the independent variables, we may construct a prediction model with more accuracy using machine learning models, which are instances of nonparametric and nonlinear models. Air pollution becomes a significant problem in the winter due to the city's low elevation and seasonal temperature inversion. Fishing, shipbuilding, textiles, pharmaceuticals, medical technology, aluminium, ferrous metals, and the state of Andhra Pradesh are all heavily concentrated in this town.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue VIII Aug 2025- Available at www.ijraset.com

You may find a 1000 MW thermal power plant on the same property. Two of the most important naturally occurring minerals, bauxite and manganese, enable heavy industrial sectors to flourish. In Andhra Pradesh, India, Visakhapatnam is among the most polluted cities based on historical AQI readings and the city's economic development. Central Pollution Control Board data made available to the public regularly is examined in this research[25-32].

Figure 2.1: An architectural prediction of air pollution based on the following;

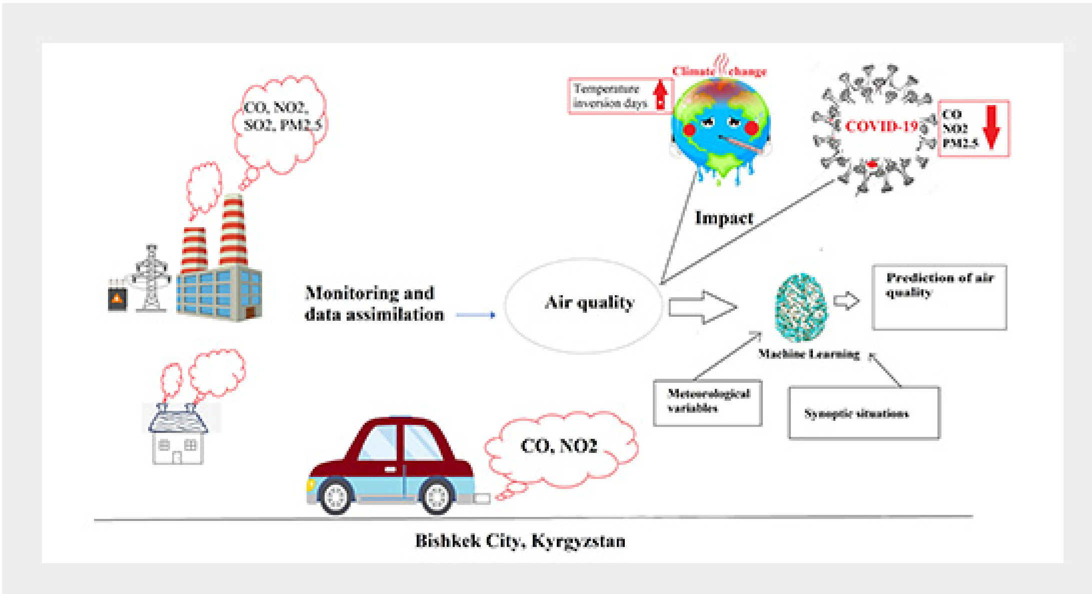


Figure 1: Air pollution prediction

II. LITERATURE SURVEY

Al Janabi and colleagues (2020) This paper describes a revolutionary intelligent computation-based air pollution prediction system. Though no specific information on RNN use was provided in this reference, the technique almost probably uses some AL or ML. A more precise approach to predicting future levels of air pollution was achieved using artificial intelligence technology. Because of the generalizability of deep learning methods like RNN, the research could not apply to a wide range of situations.

Aram and colleagues (2024) The authors forecasted air quality indicators and grades using a comparative study including different machine learning approaches. Although the research does not focus only on RNNs, it does offer comparisons to RNN-based models. It compares many machine learning models, including RNNs, and gives insight into their relative effectiveness in air quality prediction. The general focus on multiple models can diffuse attention to the specific advantages and limitations of RNNs.

Calatayud et al. (2023)This research used machine learning to forecast how electrifying cars will affect air quality in cities and the health risks that come with it. The focus on prediction models covers a broad range, including RNNs.The study connects environmental changes with prediction models, which can provide a complicated dataset suitable for RNN applications. RNN specifics are thoroughly examined since the study covers many consequences beyond prediction.

Castelli et al. (2020)forecasted the air quality in California using machine learning. Traditional machine learning seems to be the primary focus, while RNNs might or can not be employed.One advantage of using machine learning on a large and diverse dataset such as California's air quality is Because the explicit use of RNNs is unclear, the findings may be less applicable to RNN-specific research.

Chang et al. (2020)To forecast air pollution, this research developed an aggregated model based on LSTMs. Air pollution prediction using RNNs is a direct application of this work as Long Short-Term Memory (LSTM) is a kind of RNN. For time-series prediction problems like air pollution forecasting, LSTM models work well because they can grasp temporal correlations. In environments with limited resources, LSTM models may be computationally intensive and need large datasets for training, which was a drawback.

Gokul et al. (2023)This work uses AI methods to analyse air quality and estimate PM2.5 levels in Hyderabad, India, using spatiotemporal data. Although the primary focus was on artificial intelligence approaches in general, the study almost definitely employs temporal data, perhaps involving the usage of **RNNs** or LSTM models to represent the time-series nature of air quality measurements.

The spatiotemporal analysis comprehensively explains how air quality varies across time and location, improving prediction accuracy. Given that the study covers a broader range of artificial intelligence technologies, it can not focus intensely on RNNs.

Hardini et al. (2023) The writers used ensemble machine learning to forecast the air quality index (AQI). Ensemble approaches use many models to increase prediction quality. Even though the study focuses on ensemble approaches, RNNs might be used to capture temporal connections in air quality data. Ensemble approaches improve accuracy and robustness by using the capabilities of several models, including RNNs—the particular roles of RNNs in predictionare overshadowed by the current fixation on ensemble learning.

Harish Kumar et al.(2020)To predict levels of PM2.5, this research used machine learning regression models. Though the study does not expressly address RNNs, the regression models used can include modifications handling temporal data, such as LSTM models, which are a kind of RNN. Regression models, mainly when applied to time-series data, are straightforward and efficient prediction tools. The article investigates how effectively RNNs represent long-term dependence compared to other regression models.

Kumar & Pande (2023)The research forecasts urban air pollution in India using machine learning. The research will likely use several machine learning methods, but RNNs could help predict pollution levels over time.Provides a targeted focus on air pollution in Indian cities, perhaps employing RNNs to match specific temporal patterns in the data. The overall emphasis on machine learning can draw attention away from the benefits and challenges of using RNNs.

Lavanya et al. (2024) study using linked machine learning prediction algorithms and long-range radio technologies for community-level particulate matter detection via IoT. Given IoT and long-term data collection, RNNs such as LSTM models might be utilised to estimate future air quality based on current data flows. Real-time data collection enabled by IoT and machine learningwas ideal for RNNs modelling and forecasting air quality over time. The study encompasses a variety of technologies, perhaps shifting the emphasis away from the specific temporal prediction capabilities of RNNs and towards a broader focus.

Table 2.1Air pollution Prediction of comparison works

Reference	Methodology	Advantages	Limitations
Al-Janabi et al. (2020)	Took an innovative computational approach to forecasting air pollution.	Improved accuracy in predicting air pollution levels by integrating computational intelligence techniques.	Lack of specificity in applying deep learning techniques like RNN; potential limitation in scope.
Aram et al. (2024)	Compare the accuracy of several machine learning models for predicting air quality indices and grades.	Offers a comparative perspective on different machine learning models, possibly including RNNs.	The general focus on multiple models might dilute emphasis on RNN-specific advantages and challenges.
Calatayud et al. (2023)	We used machine learning to foretell how the electrification of vehicles will affect air pollution in cities and how it will affect people's health.	It connects environmental changes with predictive models and is suitable for RNN applications.	RNN specifics might not be thoroughly explored; a broader focus is on impacts beyond prediction.
Castelli et al. (2020)	I have used machine learning to forecast California's air quality.	It applies to a large and varied dataset like California's air quality.	Unclear use of RNNs, reducing relevance for RNN-specific analysis.

Chang et al. (2020)	To predict air pollution, we used an aggregated model based on long short-term memory (LSTM)	Air pollution forecasting is one example of a time-series prediction problem at LSTM, a kind of RNN excels at.	Training LSTM models requires massive datasets and is computationally demanding.
Gokul et al. (2023)	This study used artificial intelligence methods to analyse the air quality in Hyderabad City, India, and estimate PM2.5 levels at different time intervals.	It thoroughly analyses how air quality changes across different locations and times.	Broad focus on AI techniques; might not delve into RNNs specifically.
Hardini et al. (2023)	To forecast the AQI, we used ensemble machine-learning methods.	Ensemble methods improve robustness and accuracy, potentially including RNNs.	Focusing on ensemble learning might limit the emphasis on specific RNN contributions.
Harish Kumar et al. (2020)	Predicted PM2.5 values using ML regression models as the primary focus.	Regression models are adequate for prediction tasks, especially when adapted to time-series data.	It may not explore the advantages of RNNs in capturing long-term dependencies.
Kumar & Pande (2023)	They examined air pollution prediction in Indian cities using machine learning approaches.	Localised focus on Indian cities; RNNs could be used to account for unique temporal patterns in the data.	The broad focus might reduce the emphasis on RNN-specific benefits and challenges.
Lavanya et al. (2024)	She explored long-range radio technology for IoT-based particulate matter detection and prediction using machine-learning approaches.	Real-time data collection is ideal for RNNs to model and predict air quality over time.	It might not focus solely on RNNs; attention could be diffused across various technologies.

A. Feature extraction based on air pollution prediction

Li et al. (2023) this study provide a basic overview of machine learning applications that use feature extraction methods in air pollution research. It evaluates trends and research impacts without consideration of specific prediction modelsand presents a comprehensive analysis of research trends and techniques, noting differences in feature extraction methods across studies.The broad research does not further explore any feature extraction approach or its relevance in air pollution prediction.

Liu et al. (2024)This research uses machine learning techniques to foretell air quality classes using secondary data and models. This method collects data characteristics from the monitoring to improve the model's accuracy. The accuracy of air quality forecasts is enhanced by feature extraction techniques, which enhance the input data. The studydetails the specific feature extraction techniques used since it was primarily concerned with overall model performance.

Liu et al. (2022)this study investigates feature extraction in data-driven machine learning algorithms for environmental pollution. It stresses the benefits and pitfalls of various machine-learning techniques. Provides a broad understanding of how feature extraction might enhance environmental pollution machine learning models. General difficulties and improvements are highlighted instead of focusing on specific feature extraction approaches or their use in air pollution prediction.

Ma et al. (2020)assess the air quality at additional stations using a Bi-LSTM network trained via spatial transfer. Increasing prediction accuracy necessitates feature extraction from spatially transmitted data.Using bi-lSTM networks with spatially dispersed features enhances both generalisability and prediction accuracy.Rather than a detailed analysis of feature extraction approaches, the work focused on LSTM models and spatial transfer.

Ma et al. (2020)this study use XGBoost machine learning to anticipate PM2.5. Feature extraction was the selection of essential data properties to improve model performance.Good feature extraction sharpens the input data and improves prediction accuracy, enhancing XGBoost's performance.Although feature extraction was critical, the study prioritises model performance over a comprehensive assessment of feature extraction approaches.

Maltare and Vahora (2023)Predicting Ahmadabad's Air Quality Index (AQI) using characteristics extracted from many data sources wasthe focus of this work, which employs machine learning methods.Feature extraction helps identify relevant data patterns, improving AQI predictions' accuracy. The studyprovides complete knowledge of the feature extractionmethodologies employed.

Mampitiya et al.(2023)To predict future air quality, this study extracts characteristics from pertinent meteorological data using machine learning techniques.Air quality predictions become more accurate when features are extracted from meteorological data more efficiently.The focus of the study on many machine learning algorithms can limit the depth of coverage of feature extraction strategies.

Masood and Ahmad (2021) examine novel artificial intelligence methodologies, including feature extraction techniques employed in many models for air pollution forecasting. It overviews advanced feature extraction approaches and their applications in air pollution prediction. Reviewing this indicates it does not focus on specific case studies or in-depth explorations of certain feature extraction approaches.

Matthaios et al. (2024)The use of massbalance and machine learning techniques allows for predicting real-time air pollution exposure inside a vehicle. Feature extraction enhanced prediction accuracy when on-road and air quality data are included.Real-time data combined with feature extraction provides a powerful tool for predicting air pollution exposure levels.The study did not dig carefully into the specific techniques of feature extraction applied.

Table 2.2 Feature extraction methods used in air pollution prediction

Reference	Methodology	Feature Extraction Techniques	Advantages	Limitations
Li et al. (2023)	Examination of machine learning methods used in studies on air pollution.	They are generally speaking,feature extraction techniques should be included in general studies.	A comprehensive survey of research trends and advancements in feature extraction across studies.	Does not detail specific feature extraction methods or their effectiveness in prediction.
Liu et al. (2024)	Artificial intelligence techniques for air quality class prediction using secondary modelling and monitoring data.	Feature extraction from monitoring data to refine input data.	Enhances prediction accuracy by improving the quality of input data through feature extraction.	Specific feature extraction techniques are not detailed; they focus on overall model performance.
Liu et al. (2022)	Data-driven machine learning approaches for environmental pollution, including feature extraction.	General feature extraction methods are discussed in the context of environmental pollution models.	Broad understanding of how feature extraction improves machine learning models for pollution.	General discussion on challenges and gains; lacks detailed analysis of specific feature extraction methods.

Ma et al. (2020)	Predicts air quality at newlyestablished stations using a Bi-LSTM network that uses transmitted spatial characteristics.	Spatially transferred features extracted to enhance model performance.	Enhances prediction accuracy and generalizability with spatially transferred features.	Focuses on LSTM models and spatial transfer rather than detailed feature extraction methods.
Ma et al. (2020)	Predicts PM2.5 using the XGBoost machine learning approach, using features selected from relevant data.	Feature extraction involves selecting relevant data features for model performance.	It improved model performance and prediction accuracy through effective feature extraction.	It focuses on model performance and does not provide an in-depth analysis of feature extraction techniques.
Maltare & Vahora (2023)	Uses machine learning for predicting AQI in Ahmedabad city with feature extraction from various data sources.	Extraction from various data sources to identify critical patterns.	Identifies critical data patterns to enhance AQI prediction accuracy.	Limited insights into specific feature extraction methods used.
Mampitiya et al. (2023)	Feature extraction on meteorological variables is used to make air quality predictions using machine learning algorithms and meteorological data.	Feature extraction from meteorological data to improve prediction accuracy.	Effective extraction improves the accuracy of air quality predictions from meteorological data.	Focus on machine learning techniques may limit the depth of discussion on specific feature extraction methods.
Masood & Ahmad (2021)	Reviews AI techniques for air pollution forecasting, including feature extraction methods in various models.	Overview of advanced feature extraction techniques in AI-based models.	Provides a comprehensive overview of advanced feature extraction techniques and their applications.	Review format: lacks specific case studies or detailed analyses of individual feature extraction methods.
Matthaios et al. (2024)	Utilises mass-balancing and machine learning techniques to forecast in-vehicle air pollution exposure in real-time.	The extraction of features via the integration of data on roads and air quality.	Combines real-time data and feature extraction for robust prediction of air pollution exposure.	It may not delve deeply into the specifics of feature extraction techniques used.

B. Neural Networks-Based Classification for Air Pollution

Méndez et al. (2023) examine several machine learning methods, such as neural networks,to forecast air quality. It covers all the bases by outlining all the techniques used in the field. When predicting air quality, these writers are laser-focused on improving neural networks and discuss advances in neural network architecture and their implications for prediction.It primarily focuses on machine learning methods without delving into the specifics of neural network models or their relative usefulness in air pollution prediction.

Rakholia et al. (2023)conduct regional air pollution forecasts using a multi-output machine learning algorithm. The model almost probably uses neural network techniques to handle many output variables and estimate regional pollution levels. Simultaneous forecasting of multiple air pollutants, enabled by the multi-output approach, improves the model's utility for regional pollution management.

Ravindran et al. (2023) use neural network-based machine learning algorithms to forecast Visakhapatnam's air quality.The study focused on model accuracy and prediction performance.

These authors demonstrate how effectively neural networks handle complex air quality prediction situations, employing deep learning approaches to improve accuracy. Though thereare few nuances in neural network design and feature extraction approaches, overall model performance should take precedence.

Ravindra et al. (2023) employ machine learning approaches, presumably involving neural networks for classification tasks, to predict how air pollution affects health outcomes. These authors demonstrate the usefulness of neural networks in public health and environmental monitoring by combining air quality data with health impact forecasts.It is possible that the neural network architecture or feature extraction methods used for air quality prediction will not be thoroughly examined.

Sternberg et al. (2021) may use neural networks for feature extraction and prediction while investigating the variation of PM1 pollution using explainable ML approaches. It highlights how many factors impact pollution levels and how interpretable machine learning algorithms are particularly neural networks.The study might focus on explaining ability rather than the intricacies of neural network implementations and their relative effectiveness.

Suthar et al. (2024), Using machine learning techniques, including neural networks, simulate the link between air pollution, urban characteristics, and Bengaluru's land surface temperature. These authors demonstrate how to use neural networks for various environmental projections, addressing the connection between air pollution and the surface temperature of the land.The focus wasbroader, including different environmental parameters, which may limit the depth of discussion on specific neural network classification approaches.

Taheri and Razbin (2021) forecast CO_2 concentrations for indoor air quality control using machine learning, perhaps using neural networks. The technique might incorporate feature extraction from sensor data. Employs neural networks to enhance indoor air quality, demonstrating their flexibility to various air pollution conditions. Little attention was paid to outdoor air pollution and the specifics of the neural network models employed in prediction.

Tang et al. (2024)examine air quality machine learning methods, such as neural networks, and resolve neglected problems with model implementation.These authors critically review neural network applications in air quality modelling, highlighting flaws in current methodologies. Regarding reviews, it may lack comprehensive case studies and viable neural network model implementations.

Varade et al. (2023) propose a framework for IoT and machine learning-based air pollution monitoring in smart cities that may incorporate neural networks for categorisation. Uses the Internet of Things (IoT) with machine learning to improve intelligent city air quality monitoring and prediction, demonstrating the practical use of neural networks in real-time applications.Focused on IoT integration may overshadow specific neural network approaches deployed.

Veeranjaneyulu et al. (2023) apply machine learning approaches, including neural networks, to improve and optimise air quality. They demonstrate the usefulness of neural networks in enhancing and optimisingair quality prediction models. It perhaps focuses more on optimisation difficulties rather than providing a profound grasp of neural network classification algorithms.

Table 2.3: Neural Networks-Based Classification for Air Pollution

Reference	Methodology	Advantages	Limitations
Méndez et al. (2023)	A review of neural networks and other machine learning methods used for air quality prediction.	Highlights evolution and effectiveness of neural networks; discusses advancements in architectures.	Broad focus on machine learning; lacks deep analysis of neural network models and comparative performance.
Rakholia et al. (2023)	Neural network-based multi-output machine learning model for predicting regional air pollution levels.	Allows simultaneous forecasting of multiple pollutants,which is proper for comprehensive regional management.	Limited details on neural network architecture and feature extraction methods.

Ravindran et al. (2023)	They predict Visakhapatnam's air quality using machine learning models like neural networks.	Demonstrates effectiveness of neural networks; potentially uses deep learning for better accuracy.	Limited details on neural network configuration and feature extraction methods.
Ravindra et al. (2023)	Machine learning methods, perhaps including neural networks, for forecasting the effects of air pollution on health outcomes.	Integrates air quality data with health impact predictions; shows practical applications in public health.	They can not delve deeply into neural network architecture or feature extraction methods specific to forecasting.
Stirnberg et al. (2021)	Explainable machine learning approaches for analysing PM1 pollution variability may incorporate neural networks.	Focuses on interpretability; enhances understanding of feature effects on pollution levels.	Emphasis on explainability may overshadow the specifics of neural network implementations.
Suthar et al. (2024)	Machine learning, including neural networks,to forecast the temperature of the earth's surface and the correlation between it and air pollution.	Shows the use of neural networks for multi-faceted environmental predictions.	Broad focus may limit the depth of neural network classification methods.
Taheri & Razban (2021)	Machine learning for predicting CO2 concentration for indoor air quality may include neural networks.	Versatile application of neural networks in indoor air quality control.	There is a limited focus on outdoor air pollution; the specifics of neural network models are not detailed.
Tang et al. (2024)	Analysing air quality models using machine learning, including neural networks, addresses implementation issues.	Provides critical review; identifies gaps and challenges in neural network applications.	It can not offer detailed case studies or practical implementations of neural network models.
Varade et al. (2023)	Internet of Things (IoT) and machine learning framework for evaluating intelligent city air pollution; includes neural networks.	It uses machine learning and the Internet of Things to monitor air quality in real-time.	The focus on IoT integration may overshadow the specific neural network techniques used.
Veeranjaneyulu et al. (2023)	Neural networks and other machine learning are used for air quality improvement and optimisation.	Demonstrates practical applications of neural networks for improving prediction models.	It can focus more on optimisation than detailed insights into neural network methodologies.

III. EXISTING METHODOLOGY

A. RNN(Recurrent Neural Networks)

The present step in a recurrent neural network (RNN) takes data from the prior step and uses it to influence its own decisions. Each input and output of a conventional neural network operates independently. However, the ability to recall the prior words becomes vital because predicting the next word in a phrase depends on them. Using a Hidden Layer, RNN emerged to address this problem. The most crucial aspect of RNN is its hidden state, which stores particular sequence information. Another name for this state is Memory State since it keeps track of what the network has already received. It uses the same method on each input or hidden layer to create the output, using the same parameters. Unlike other neural networks, this reduces parameter complexity.

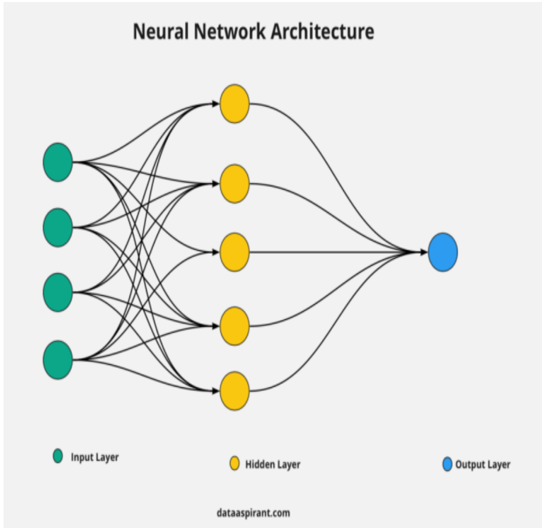


Figure 2: RNN Architecture

Works for RNN

Convolutional neural networks (CNNs) use recurrent neural networks to learn from training data. Their "memory" makes them unique; it allows them to draw on past inputs to shape the present input and output. The underlying assumption of recurrent neural networks is that inputs and outputs are independent, in contrast to the more conventional deep neural networks. Unidirectional recurrent neural networks cannot consider future events when making predictions, even if these events might help determine the outcome of a particular sequence.

Consider the term "feeling under the weather" as a tool to comprehend RNNs better. If you want to use the idiom correctly, you must say it in that exact sequence. Determining the following word in the sequence relies on recurrent networks analysing the location of each word in the phrase.

Recurrent networks vary in that each layer shares parameters with the others. Recurrent neural networks maintain a constant weight parameter throughout all layers, unlike feedforward networks that experience weight changes at each node. Reinforcement learning, however, is achieved by adjusting these weights throughout the backpropagation and gradient descent procedures.

Recurrent neural networks use BPTT methods to find the gradients, which are unique to sequence data and vary slightly from conventional backpropagation. Like classic backpropagation, BPTT relies on the model learning process of feeding mistakes back into the input layer from the output layer. Through these computations, we can fine-tune and calibrate the model's parameters. Unlike conventional methods, BPTT adds up mistakes at each time step. This contrasts feedforward networks, which do not share parameters between layers.

Problems with disappearing and inflated gradients are shared throughout this procedure for RNNs. These gradients' magnitude, defined as the loss function's slope across the error curve, characterise these problems. If the gradient is too tiny, it will keep becoming smaller and smaller until it changes the weight parameters to zero. The method's ability to learn is now gone. One way to address these concerns is by simplifying the RNN model by lowering the number of hidden layers in the neural network.

B. LSTM(Long Short-Term Memory)

The memory unit, or Long Short-Term Memory (LSTM) unit, is the building block of recurrent neural networks. In the LSTM unit, four neural networks are feedforward. There are two layers to each of these neural networks: input and output. Neural connections between input and output nodes are present in these neural networks. Consequently, the LSTM unit is made up of four interconnected layers. Data selection is handled by three of the four feedforward neural networks. Remember, input and output are the three gates that make them up. There are three typical procedures in memory management that these three gates facilitate: removing data from memory (the forget gate), adding data to memory (the input gate), and using data already in memory. The candidate memory, the fourth neural network, creates and stores new candidate information.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue VIII Aug 2025- Available at www.ijraset.com

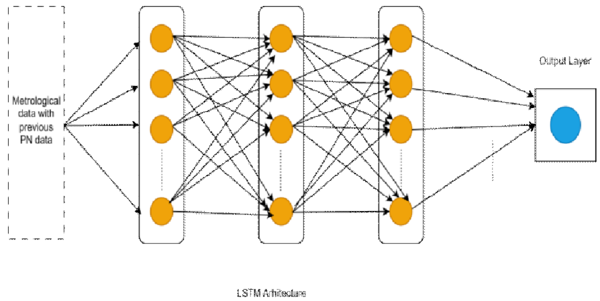


Figure 3: LSTM Architecture

1) Types of RNN

Based on the amount of inputs and outputs, four distinct kinds of RNNs may be distinguished.

- One to One

One sort of RNN is the Vanilla Neural Network, which acts like any other leading neural network. This neural network takes in one data point and produces one result.

- One To Many

The characteristics of this RNN type are a single input and several outputs. Image captioning is a widely used network applicationinvolving taking pictures and predicting multi-word statements.

- Many to One

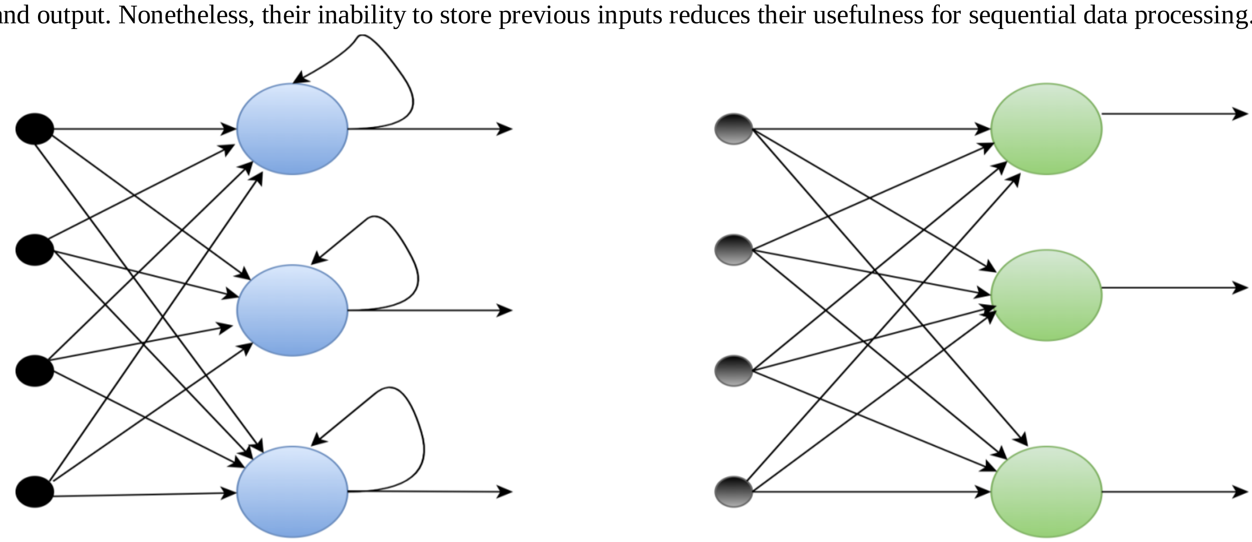
A single output is produced from a network that receives several inputs at different stages. Situations like sentiment analysis benefit from this kind of network. Given many words as input, how can we predict the text's mood as an output?

- Many to Many

Numerous input and problem-specific outputs characterise this kind of neural network. Language translation might be a part of this issue. When we translate from one language to another, we usually end up with several new terms in the target language.

2) RNN differs from Feedforward Neural Network

Feedforward neural networks are artificial neural networks that do not use node loops. Since all input is transmitted forward, this kind of neural network is called a multi-layer neural network. There is no need for hidden layers in a feedforward neural network; data flows directly from input to output. These networks are appropriate for image classification applications that need independent input and output. Nonetheless, their inability to store previous inputs reduces their usefulness for sequential data processing.



a)Recurrent neural Network

b)Feed Forward Neural Networks

Figure 4 a) Recurrent neural network and b) Feedforward neural network

C. Feedforward neural network

An FNN is a kind of artificial neural network in which the connections between the nodes do not create cycles. This is one way in which it deviates from RNNs. An input layer, a hidden layer (or layers), and an output layer comprise the network. When data goes straight from input to output, we say it is feedforward.

Structure of a Feedforward Neural Network

- Input Layer: Neurons in the input layer take in information from the outside world. In the input layer, every neuron stands for a different attribute of the data that is being inputted.
- Hidden Layers: There is a latent layer or layers between the input and output layers. Only at these levels can one hope to learn the intricate patterns within the data. Every neuron in a buried layer receives a weighted sum of inputs before being activated non-linearly.
- Output Layer: The network's result is controlled by the output layer. The number of outputs in a regression issue correlates to the number of neurons in this layer, whereas in a classification challenge, it corresponds to the number of classes.

D. GBM in Machine Learning

Gradient boosting machine (GBM) is one of the most used ML forward learning ensemble methods. Classification and regression prediction models may be effectively created utilising this strategy.Using GBM, we may combine many weak prediction models, such as decision trees, into a single predictive model. Gradient-boosted trees are the resultant method if a decision tree shows as a poor learner.

Machine learning is one of the most used tools for developing prediction models for complex classification and regression issues. The gradient boosting machine is among the most effective boosting methods. Though machine learning uses so many techniques, boosting algorithms has grown very popular within the worldwide machine learning community. Inspired by ensemble learning, the boosting method combines many rudimentary models—weak learners or base estimators—to get the output. In machine learning, GBM is also a collective approach that transforms weak learners into strong learners.

The course "GBM in Machine Learning" will cover various topics, like gradient machine learning techniques, GBM's history, how it works, a few terms related to GBM, and more. Before beginning, however, first, grasp the boosting idea and many boosting techniques in machine learning.

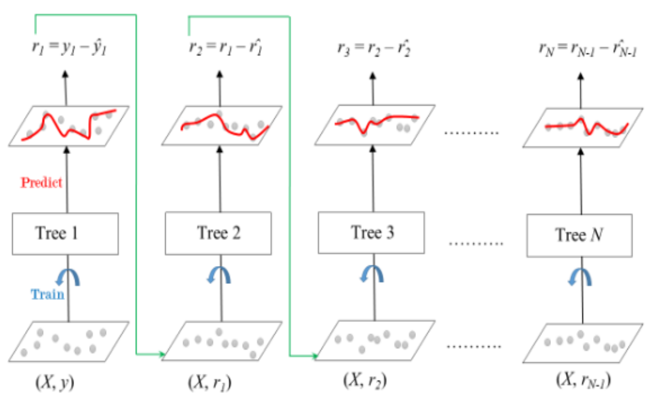


Figure 5:GBM architecture Machine learning

1) Steps in Boosting Algorithms:

There are a few essential steps in boosting the algorithm as follows:

- Create a dataset with many data points and examine their differences. For the time being, treat each piece of data equally.
- Let the model use this weight as input.
- Highlight the incorrectly categorised data elements.
- Step 4's data point weight should be increased.
- If you produce satisfactory results, terminate the method; otherwise, try steps 2 and 3 again.

2) Boosting Algorithms in Machine Learning

There are primarily **fourmain** boosting algorithms in machine learning. These are as follows:

- Gradient Boosting Machine (GBM)
- Extreme Gradient Boosting Machine (XGBM)
- Light GBM
- CatBoost

3) Advantages of Boosting Algorithms:

- Boosting approaches are based on ensemble learning, enabling a model to produce an unchanging, more accurate prediction.
- Boosting approaches provide several hyperparameter tweaking options and may maximise different loss functions, making them far more versatile than other algorithms.
- It is suited for numerical and categorical variables; therefore, data preparation is not required.
- It automatically handles missing data and does not need imputation of missing values in the dataset.

4) Disadvantages of Boosting Algorithms:

Below are a few disadvantages of boosting algorithms:

- Boosting algorithms may overfit and overemphasise outliers.
- Gradient boosting is computationally expensive since it relies on several trees and focuses on reducing errors.
- It is a memory-intensive, time-consuming approach.
- It is less interpretative, although specific tools can manage this.

IV. DISCUSSIONS

Dealing with environmental and public health concerns heavily depends on air pollution forecasts. Accurate pollution level estimates facilitate urban planning, regulatory compliance, timely interventions, and policy guidance.Because it can process complicated data and see patterns that more conventional approaches might overlook, machine learning (ML) has become an invaluable resource for this sector. Several machine learning techniques, such as random forests, linear regression, support vector machines (SVMs), and deep learning models, have been used to forecast air quality.

There is a broad spectrum of models here, from basic regression techniques to sophisticated deep neural networks that can spot patterns in data. Variations in monitoring techniques, area coverage, and temporal precision may cause air pollution numbers to be inconsistent. Accurate forecasts rely on data quality and consistency being ensured. Among other applications, models are used for real-time monitoring, forecasting, and long-term trend analysis. A model's effectiveness is affected by its unique application. Including real-time data in prediction algorithms will improve their accuracy and responsiveness. Studies on approaches for effectively integrating and interpreting real-time data are ongoing. Combining meteorological data with ground-based sensors and satellite photos, among other sources, enhances forecast models. Future research should focus on constructing interpretable decision-making models scalable to large datasets. Model adoption and utilisation depend on stakeholders' ability to comprehend and trust them.

V. CONCLUSIONS

The survey paper focuses on the current state of research, challenges, and future possibilities for air pollution prediction using machine learning. Machine learning has significantly improved air pollution prediction by combining various methodologies and data sources. Although it remains challenging due to data unpredictability and various component integration, emerging technologies promise to increase real-time management and prediction accuracy. Future research should address these difficulties and explore innovative approaches to improving air quality management and prediction. Machine learning has improved air pollution prediction by providing sophisticated methods for modelling and forecasting pollution levels. Despite the challenges, ongoing improvements in algorithms, data integration, and new technologies promise to improve air quality management and prediction accuracy. Future research will push the boundaries of what is possible, aiming to develop robust and valuable models for real-world applications.

REFERENCES

[1] Al-Janabi, S., Mohammad, M., & Al-Sultan, A. (2020). A new method for prediction of air pollution based on intelligent computation. Soft Computing, 24(1), 661-680. https://doi.org/10.1007/s00500-019-03972-6

[2] Aram, S. A., Nketiah, E. A., Saalidong, B. M., Wang, H., Afitiri, A. R., Akoto, A. B., & Lartey, P. O. (2024). Machine learning-based prediction of air quality index and air quality grade: A comparative analysis. International Journal of Environmental Science and Technology, 21(2), 1345-1360. https://doi.org/10.1007/s13762-023-04523-4

[3] Calatayud, V., Diéguez, J. J., Agathokleous, E., & Sicard, P. (2023). Machine learning model to predict vehicle electrification impacts on urban air quality and related human health effects. Environmental Research, 228, 115835.

[4] Castelli, M., Clemente, F. M., Popovič, A., Silva, S., & Vanneschi, L. (2020). A machine learning approach to predict air quality in California. Complexity, 2020(1), 8049504.

[5] Chang, Y. S., Chiao, H. T., Abimannan, S., Huang, Y. P., Tsai, Y. T., & Lin, K. M. (2020). An LSTM-based aggregated model for air pollution forecasting. Atmospheric Pollution Research, 11(8), 1451-1463.

[6] Gokul, P. R., Mathew, A., Bhosale, A., & Nair, A. T. (2023). Using artificial intelligence techniques, spatiotemporal air quality analysis, and PM2. 5 predictions over Hyderabad City, India. Ecological Informatics, 76, 102067.

[7] Hardini, M., Sunarjo, R. A., Asfi, M., Chakim, M. H. R., & Sanjaya, Y. P. A. (2023). They are predicting air quality index using ensemble machine learning. ADI Journal on Recent Innovation, 5(1Sp), 78-86. https://doi.org/10.46799/adijri.v5i1sp.290.

[8] Harishkumar, K. S., Yogesh, K. M., & Gad, I. (2020). Forecasting air pollution particulate matter (PM2. 5) using machine learning regression models. Procedia Computer Science, 171, 2057-2066.

[9] Kumar, K., & Pande, B. P. (2023). Air pollution prediction with machine learning: A case study of Indian cities. International Journal of Environmental Science and Technology, 20(5), 5333-5348. https://doi.org/10.1007/s13762-023-04827-4.

[10] Lavanya, P., Reddy, I. S., & Selvakumar, V. (2024). Long-range radio technology implementation on the Internet of Things to detect particulate matter at the community level and predict using a machine learning approach. Engineered Science, 29, 1119. https://doi.org/10.30919/es8d1119.

[11] Li, Y., Sha, Z., Tang, A., Goulding, K., & Liu, X. (2023). The application of machine learning to air pollution research: A bibliometric analysis. Ecotoxicology and Environmental Safety, 257, 114911. https://doi.org/10.1016/j.ecoenv.2023.114911.

[12] Liu, Q., Cui, B., & Liu, Z. (2024). Air quality class prediction using machine learning methods based on monitoring data and secondary modelling. Atmosphere, 15(5), 553. https://doi.org/10.3390/atmos15050553.

[13] Liu, X., Lu, D., Zhang, A., Liu, Q., & Jiang, G. (2022). Data-driven machine learning in environmental pollution: gains and problems. Environmental science & technology, 56(4), 2124-2133.

[14] Ma, J., Li, Z., Cheng, J. C., Ding, Y., Lin, C., & Xu, Z. (2020). Air quality prediction at new stations using spatially transferred bi-directional long short-term memory network. Science of The Total Environment, 705, 135771. https://doi.org/10.1016/j.scitotenv.2019.135771.

[15] Ma, J., Yu, Z., Qu, Y., Xu, J., & Cao, Y. (2020). Application of the XGBoost machine learning method in PM2.5 prediction: A case study of Shanghai. Aerosol and Air Quality Research, 20(1), 128-138. https://doi.org/10.4209/aaqr.2019.05.0208.

[16] Maltare, N. N., & Vahora, S. (2023). Air Quality Index prediction using machine learning for Ahmedabad city. Digital Chemical Engineering, 7, 100093. https://doi.org/10.1016/j.diche.2023.100093.

[17] Mampitiya, L., Rathnayake, N., Leon, L. P., Mandala, V., Azamathulla, H. M., Shelton, S., ... & Rathnayake, U. (2023). Machine learning techniques to predict the air quality using meteorological data in two urban areas in Sri Lanka. Environments, 10(8), 141.

[18] Mampitiya, L., Rathnayake, N., Leon, L. P., Mandala, V., Azamathulla, H. M., Shelton, S., ... & Rathnayake, U. (2023). Machine learning techniques to predict the air quality using meteorological data in two urban areas in Sri Lanka. Environments, 10(8), 141.

[19] Masood, A., & Ahmad, K. (2021). A review on emerging artificial intelligence (AI) techniques for air pollution forecasting: Fundamentals, application and performance. Journal of Cleaner Production, 322, 129072. https://doi.org/10.1016/j.jclepro.2021.129072

[20] Matthaios, V. N., Knibbs, L. D., Kramer, L. J., Crilley, L. R., & Bloss, W. J. (2024). Predicting real-time within-vehicle air pollution exposure with mass-balance and machine learning approaches using on-road and air quality data. Atmospheric Environment, 318, 120233. https://doi.org/10.1016/j.atmosenv.2024.120233

[21] Méndez, M., Merayo, M. G., & Núñez, M. (2023). Machine learning algorithms to forecast air quality: a survey. Artificial Intelligence Review, 56(9), 10031-10066.

[22] Rakholia, R., Le, Q., Ho, B. Q., Vu, K., & Carbajo, R. S. (2023). Multi-output machine learning model for regional air pollution forecasting in Ho Chi Minh City, Vietnam. Environment International, 173, 107848. https://doi.org/10.1016/j.envint.2023.107848.

[23] Ravindran, G., Hayder, G., Kanagarathinam, K., Alagumalai, A., & Sonne, C. (2023). Air quality prediction by machine learning models: A predictive study on the Indian coastal city of Visakhapatnam. Chemosphere, 338, 139518. https://doi.org/10.1016/j.chemosphere.2023.139518.

[24] Ravindra, K., Bahadur, S. S., Katoch, V., Bhardwaj, S., Kaur-Sidhu, M., Gupta, M., & Mor, S. (2023). Application of machine learning approaches to predict the impact of ambient air pollution on outpatient visits for acute respiratory infections. Science of The Total Environment, 858, 159509. https://doi.org/10.1016/j.scitotenv.2023.159509.

[25] Stirnberg, R., Cermak, J., Kotthaus, S., Haeffelin, M., Andersen, H., Fuchs, J., ... & Favez, O. (2021). Meteorology-driven variability of air pollution (PM 1) revealed with explainable machine learning. Atmospheric Chemistry and Physics, 21(5), 3919-3948.

[26] Suthar, G., Kaul, N., Khandelwal, S., & Singh, S. (2024). Predicting land surface temperature and examining its relationship with air pollution and urban parameters in Bengaluru: A machine learning approach. Urban Climate, 53, 101830. https://doi.org/10.1016/j.uclim.2024.101830.

[27] Taheri, S., & Razban, A. (2021). Learning-based CO2 concentration prediction: Application to indoor air quality control using demand-controlled ventilation. Building and Environment, 205, 108164. https://doi.org/10.1016/j.buildenv.2021.108164

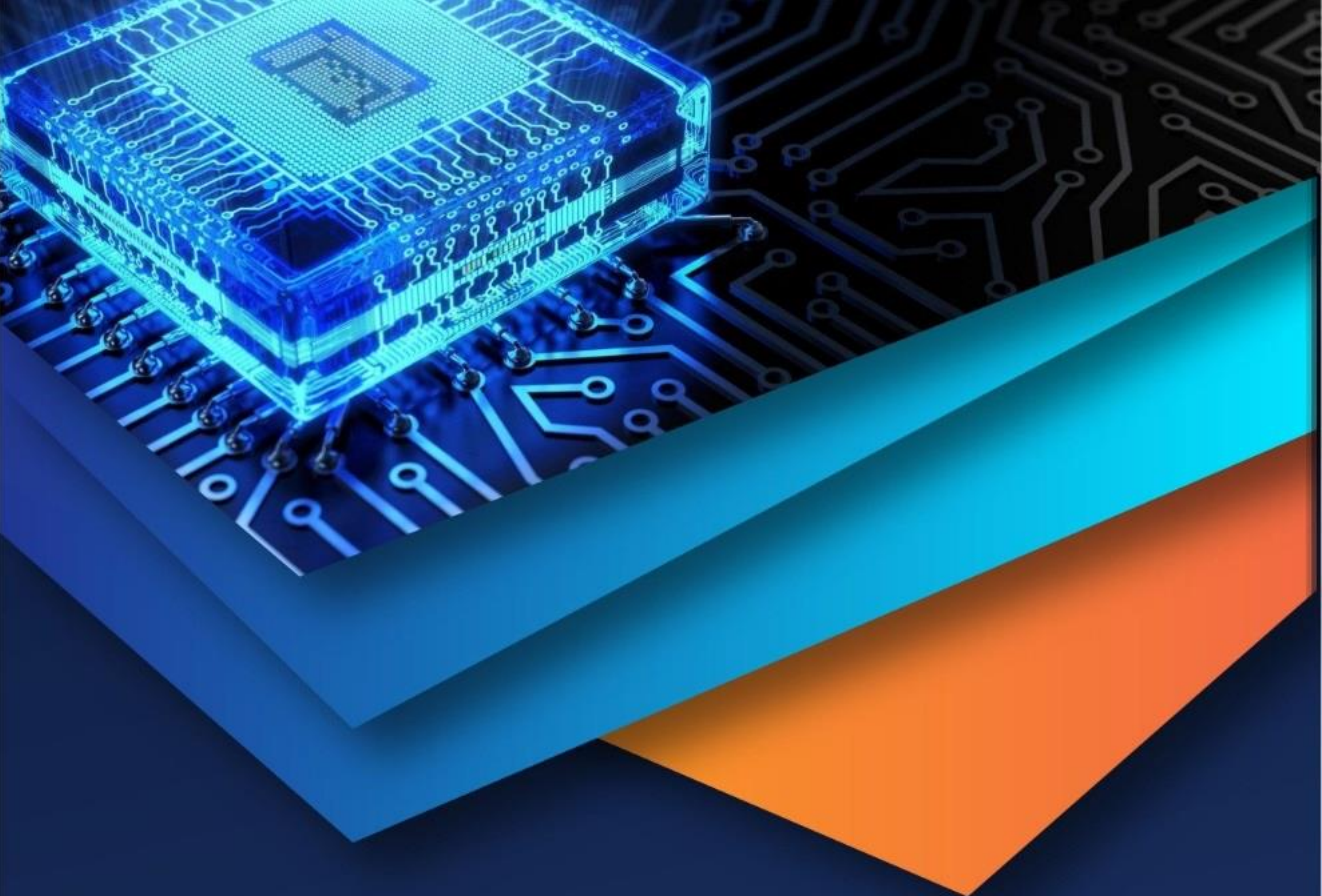
[28] Tang, D., Zhan, Y., & Yang, F. (2024). A review of machine learning for modelling air quality: Overlooked but important issues. Atmospheric Research, 107261. https://doi.org/10.1016/j.atmosres.2024.107261

[29] Varade, H. P., Bhangale, S. C., Thorat, S. R., Khatkale, P. B., Sharma, S. K., & William, P. (2023, May). The framework of air pollution assessment in smart cities using IoT with machine learning approach. In 2023 2nd International Conference on Applied Artificial Intelligence and Computing (ICAAIC) (pp. 1436-1441). IEEE. https://doi.org/10.1109/ICAAIC58029.2023.10217567.

[30] Veeranjaneyulu, R., Boopathi, S., Kumari, R. K., Vidyarthi, A., Isaac, J. S., & Jaiganesh, V. (2023, May). Air quality improvement and optimisation using machine learning techniques. In 2023 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI) (pp. 1-6). IEEE. https://doi.org/10.1109/ACCAI55650.2023.10157934.

[31] Wang, S., McGibbon, J., & Zhang, Y. (2024). Predicting high-resolution air quality using machine learning: Integration of large eddy simulation and urban morphology data. Environmental Pollution, 344, 123371. https://doi.org/10.1016/j.envpol.2024.123371.

[32] William, P., Paithankar, D. N., Yawalkar, P. M., Korde, S. K., Rajendra, A., & Raksha, D. S. (2023, January). Divination of air quality assessment using ensembling machine learning approach. In 2023 International Conference on Artificial Intelligence and Knowledge Discovery in Concurrent Engineering (ICECONF) (pp. 1-10). IEEE.



10.22214/IJRASET

INDEX COPERNICUS

45.98

IMPACT FACTOR:
7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR:
7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY