A COMPUTATIONAL INTELLIGENCE PARADIGM FOR PRIOR AIR QUALITY PREDICTION

Dr.A John Clement Sunder
Dept of CSE
Bharath Inistutute of Higher Education and Research
Chennai, India
johnclementawork@gmail.com

R.Charan
Dept of CSE
Bharath Instute of Higher Education and Research)
Chennai,India
ravineni004@gmail.com

C.G.Rajasekhar Reddy
Dept of CSE
Bharath Instute of Higher Education and Research
Chennai,India
cgrajasekharreddy7@gmail.com

J.Ramakrishna
Dept of CSE
Bharath Instute of Higher Education and Research
Chennai,India
ramakrishnajagathi4@gmail.com

R.Sujay Varma
Dept of CSE
Bharath Instute of Higher Education and Research
Chennai,India
sujayvarmarudraraju12345@gmail.com

***Abstract*— Air pollution has become one of the most serious environmental challenges affecting human health and ecological balance across the world. Rapid urbanization, industrial activities, and increasing vehicular emissions have significantly contributed to the deterioration of air quality in many cities. Continuous exposure to polluted air can lead to severe health problems such as respiratory diseases, cardiovascular disorders, and reduced life expectancy. Therefore, predicting air pollution levels in advance is essential for enabling authorities and individuals to take preventive actions and reduce the harmful effects of poor air quality. This study presents a machine learning–based approach for predicting the Air Quality Index (AQI) using historical environmental data. The system analyzes various pollution parameters including PM2.5, PM10, nitrogen dioxide (NO_2), carbon monoxide (CO), sulfur dioxide (SO_2), temperature, and humidity. The dataset is preprocessed through data cleaning, normalization, and feature selection to improve model performance. Several machine learning algorithms such as Linear Regression, Random Forest, and XGBoost are applied to learn patterns from historical data and generate accurate AQI predictions. The performance of the proposed models is evaluated using standard evaluation metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and R^2 score. Experimental results demonstrate that machine learning models can effectively capture relationships between environmental factors and air pollution levels, providing reliable forecasts of future air quality conditions. The developed system can assist environmental authorities and the general public by providing early warnings of pollution levels, thereby supporting better environmental monitoring and decision-making.**

INDEX TERMS — Air Quality Prediction, Machine Learning, Air Quality Index (AQI), Environmental Monitoring, Random Forest, XGBoost, Linear Regression, Pollution Forecasting, PM2.5 and PM10 Analysis, Data Preprocessing, Environmental Data Analysis

I.INTRODUCTION

Air pollution has become one of the most critical environmental problems affecting human health and ecological balance worldwide. Rapid industrial growth, urbanization, and increasing vehicular emissions have significantly contributed to the deterioration of air quality in many cities. Polluted air contains harmful substances such as particulate matter, nitrogen oxides, carbon monoxide, and sulfur dioxide, which can cause severe respiratory and cardiovascular diseases.

Air quality monitoring has become essential for understanding pollution levels and protecting public health. Governments and environmental agencies install monitoring stations to measure pollutants and calculate the Air Quality Index (AQI). The AQI is an important indicator that reflects the level of air pollution and its potential impact on human health. Traditional air monitoring systems mainly focus on measuring the current level of pollutants in the atmosphere. Although these systems provide valuable real-time information, they are not capable of accurately predicting future air quality conditions. Without forecasting capabilities, it becomes difficult for authorities and citizens to take preventive actions in advance.

With the rapid growth of data science and artificial intelligence, Machine Learning has emerged as a powerful tool for analyzing large environmental datasets. Machine learning algorithms can identify hidden patterns in historical data and use them to predict future outcomes. This capability makes machine learning highly suitable for environmental prediction problems such as air quality forecasting. Air pollution prediction involves analyzing various environmental factors that influence pollutant concentration. These factors include meteorological parameters such as temperature, humidity, wind speed, and atmospheric pressure. In addition, pollutants like PM2.5, PM10, nitrogen dioxide (NO_2), carbon monoxide (CO), and sulfur dioxide (SO_2) play a major role in determining overall air quality.

Particulate matter such as PM2.5 and PM10 are among the most harmful pollutants affecting human health. These microscopic particles can penetrate deep into the lungs and bloodstream, causing respiratory illnesses and heart problems. Monitoring and predicting particulate matter concentration is therefore essential for maintaining public health safety.

Machine learning techniques enable researchers to analyze large volumes of historical air quality data collected from monitoring stations. By training models on past pollution data, it is possible to identify relationships between different environmental variables and pollutant concentrations. These relationships help build predictive models capable of estimating future air pollution levels. Data preprocessing plays a critical role in building accurate machine learning models. Environmental datasets often contain missing values, noise, and inconsistent measurements. Proper preprocessing techniques such as data cleaning, normalization, and feature selection are necessary to improve the reliability and accuracy of the predictive models.

Feature selection is another important step in air quality prediction. Not all environmental parameters contribute equally to pollution levels. Identifying the most relevant features helps reduce model complexity and improves prediction performance. This process ensures that the machine learning models focus on the most influential variables. Several machine learning algorithms have been successfully applied to air quality prediction problems. Regression algorithms are particularly suitable because air quality prediction involves estimating continuous values such as AQI. These algorithms learn patterns from historical data and generate predictions for future conditions.

Linear Regression is one of the most commonly used algorithms for predicting numerical values. It establishes a linear relationship between input features and the target variable. Although simple, it provides a good baseline model for understanding the relationship between environmental parameters and AQI levels. Random Forest is another widely used machine learning algorithm for environmental prediction tasks. It is an ensemble learning technique that builds multiple decision trees and combines their outputs to improve prediction accuracy. Random Forest models are known for handling complex data relationships and reducing overfitting.

XGBoost (Extreme Gradient Boosting) is a powerful boosting algorithm that has gained popularity in many data science applications. It improves prediction performance by sequentially correcting the errors of previous models. XGBoost is highly efficient and often provides better accuracy compared to traditional algorithms. Comparing multiple machine learning models helps identify the most suitable algorithm for air quality prediction. Different models may perform differently depending on the dataset characteristics. Evaluating their performance ensures that the most accurate and reliable model is selected for forecasting air pollution levels.

Model evaluation is an essential part of machine learning system development. Various performance metrics are used to assess the accuracy of predictive models. Common evaluation metrics for regression problems include Mean Absolute Error (MAE), Mean Squared Error (MSE), and the coefficient of determination (R^2 score).

The Mean Absolute Error measures the average difference between predicted and actual AQI values. A lower MAE value indicates better prediction accuracy. Similarly, Mean Squared Error measures the average squared difference between predicted and actual values, emphasizing larger errors. The R^2 score is another important metric that indicates how well the model explains the variation in the target variable. An R^2 value closer to 1 indicates a better model fit. Using these evaluation metrics helps ensure that the developed machine learning system produces reliable predictions.

Accurate air quality prediction systems can provide significant benefits for society. Early warnings about high pollution levels allow governments to implement control measures such as traffic restrictions or industrial emission reductions. These preventive actions can reduce health risks for the population. In addition, air quality forecasting systems can help individuals plan their daily activities more safely. People with respiratory problems or other health conditions can take precautions when pollution levels are expected to rise. This improves public awareness and promotes healthier living conditions.

In conclusion, the integration of machine learning techniques with environmental monitoring data provides an effective solution for air quality prediction. By analyzing historical pollution patterns and meteorological factors, machine learning models can forecast future AQI levels with improved accuracy. Such systems play an important role in supporting environmental management and protecting public health.

II. RELATED WORK

Air quality prediction has been widely studied in recent years due to the increasing impact of air pollution on human health and the environment. Researchers have explored various statistical, machine learning, and deep learning approaches to forecast pollutant levels such as PM2.5, PM10, NO_2, CO, and SO_2. These studies aim to provide accurate predictions that can help governments and citizens take preventive actions.

Earlier research mainly focused on traditional statistical models such as autoregressive integrated moving average (ARIMA) and linear regression for air pollution forecasting. Although these models can analyze historical patterns, they often struggle to capture complex nonlinear relationships present in environmental data. With the advancement of machine learning techniques, researchers began using algorithms such as Support Vector Machines (SVM), Random Forest, and Gradient Boosting for air quality prediction. These models can analyze large datasets and identify hidden relationships between environmental parameters and pollutant concentrations.

A study evaluated several machine learning algorithms for predicting PM2.5 concentrations and found that ensemble models such as Random Forest and Gradient Boosting performed well in forecasting pollution levels. The study also highlighted the importance of feature selection and data preprocessing for improving model accuracy.

Another research compared multiple machine learning methods including decision trees, neural networks, and random forests for air pollution prediction. The results showed that Random Forest achieved high prediction accuracy and was capable of handling complex relationships among environmental variables. Researchers have also applied deep learning techniques for air quality forecasting. Deep learning models can capture nonlinear relationships and temporal patterns more effectively than traditional machine learning methods, leading to improved prediction performance.

Long Short-Term Memory (LSTM) networks have been widely used for air pollution prediction because they can learn time-series patterns in environmental data. Studies have shown that LSTM models can accurately predict PM2.5 concentrations by capturing long-term dependencies in historical pollution data. A comparative study involving multiple models such as Random Forest, AdaBoost, XGBoost, and LSTM demonstrated that deep learning models often achieve better performance when dealing with complex temporal data. The research concluded that hybrid models combining different algorithms can further improve prediction accuracy.

Some researchers have developed hybrid models combining convolutional neural networks (CNN) and LSTM networks for air quality prediction. The CNN component extracts spatial features from pollution data, while the LSTM component captures time-dependent patterns, resulting in improved forecasting performance. Spatiotemporal modeling has also been explored for air pollution forecasting. These models consider both spatial relationships between monitoring stations and temporal patterns in pollutant data, providing a more comprehensive understanding of pollution dynamics.

A multimodal deep learning framework was proposed to analyze different environmental factors simultaneously. This model combined meteorological data, pollutant concentrations, and geographic features to improve the accuracy of long-term air quality predictions. Another study introduced convolutional LSTM networks for large-scale air quality forecasting. The model used weather data and pollution measurements to predict concentrations of pollutants such as NO_2, PM2.5, and PM10 several days in advance. Researchers have also emphasized the importance of meteorological variables such as temperature, humidity, and wind speed in air quality prediction models. These environmental factors significantly influence pollutant dispersion and concentration levels.

In recent years, ensemble learning techniques such as XGBoost and LightGBM have gained popularity in air quality prediction tasks. These algorithms improve prediction accuracy by combining the outputs of multiple decision trees and optimizing model performance.Several studies have shown that machine learning models trained on large datasets collected from monitoring stations can provide reliable air quality forecasts. These systems can help environmental authorities monitor pollution trends and take timely actions to control emissions.

Researchers have also explored the use of feature selection methods to identify the most influential parameters affecting air quality. Selecting relevant features helps reduce model complexity and improves prediction efficiency.

Recent research has focused on integrating physical environmental knowledge with machine learning models to improve prediction accuracy. These hybrid approaches combine data-driven learning with environmental physics to better understand pollutant movement and dispersion. For example, physics-guided neural networks have been proposed to incorporate wind direction and topographical factors into air quality prediction models. Such approaches significantly improve forecasting performance compared to traditional methods.

Many studies have also evaluated model performance using metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and R^2 score. These metrics help researchers compare different algorithms and identify the most accurate models. Despite significant progress in air quality forecasting, several challenges remain. Air pollution data often contain missing values, noise, and irregular patterns, which can affect prediction accuracy.

In addition, air pollution is influenced by multiple dynamic factors such as industrial emissions, traffic patterns, weather conditions, and geographical features. Accurately modeling these interactions remains a challenging task for researchers. Therefore, continuous research is being conducted to develop more advanced machine learning and deep learning models for air quality prediction. These models aim to improve forecasting accuracy and support better environmental management and public health protection.

Recent studies have also explored the use of Support Vector Machines (SVM) for predicting air pollution levels. SVM models are effective in handling nonlinear relationships between environmental variables and pollutant concentrations. Researchers found that SVM can provide reliable predictions when trained with well-processed environmental datasets. Another research focused on the use of Artificial Neural Networks (ANN) for air quality prediction. ANN models are capable of learning complex relationships in large datasets through multiple hidden layers. These models have shown good performance in predicting pollutant levels such as PM2.5 and PM10.

Some studies have used hybrid machine learning models that combine multiple algorithms to improve prediction accuracy. For example, combining Random Forest with neural networks can help capture both nonlinear relationships and feature importance in air quality datasets. Researchers have also investigated the role of feature engineering in improving air quality prediction models. Feature engineering involves creating new variables from existing data to enhance model performance. Proper feature engineering can significantly increase prediction accuracy.

Several studies have used Internet of Things (IoT) sensors to collect real-time environmental data for air quality prediction. IoT devices can continuously monitor pollution levels and send data to cloud servers where machine learning models analyze the information and generate predictions. Cloud computing platforms have also been used to process large volumes of environmental data for air pollution forecasting. These systems allow researchers to train complex machine learning models efficiently and deploy them for real-time prediction applications.

Another area of research focuses on spatial air quality prediction using geographical data. By analyzing pollution data from multiple monitoring stations, machine learning models can predict air quality in areas where monitoring stations are not available. Researchers have also applied transfer learning techniques for air quality prediction. Transfer learning allows models trained in one location to be adapted for predicting air pollution in another location with limited data.

Some studies have investigated the use of deep reinforcement learning for environmental monitoring and pollution control. These models can learn optimal strategies to reduce pollution levels by analyzing environmental data and emission patterns. Overall, previous research demonstrates that machine learning and deep learning techniques provide powerful tools for air quality prediction. However, continuous improvements in data quality, feature selection, and model design are necessary to develop more accurate and reliable air pollution forecasting systems.

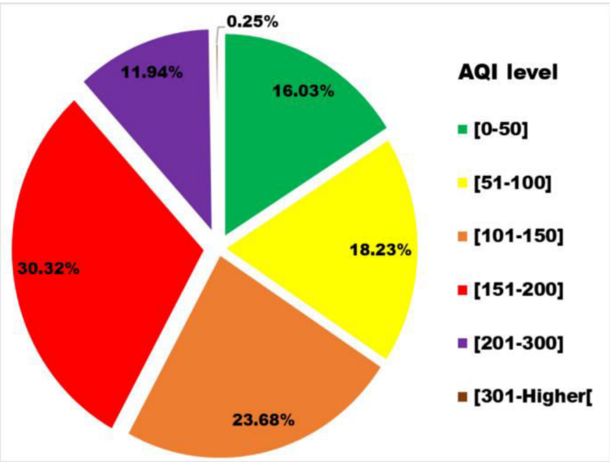

Fig 1: Air Quality data classes 1

III. *PROPOSED METHOD*

The proposed system focuses on developing an efficient and accurate air quality prediction model using machine learning techniques. The system is designed to analyze historical environmental and pollution data to forecast future air quality levels. The proposed approach aims to improve prediction accuracy by utilizing machine learning algorithms that can identify hidden patterns in environmental datasets. The system consists of several stages including data collection, data preprocessing, model training using machine learning algorithms, and air quality prediction. The architecture is designed to effectively learn patterns from historical data and generate reliable predictions that can assist environmental authorities and the public in taking preventive actions against air pollution.

A. Data Collection

The first step in the proposed system is the collection of historical air quality data. The dataset includes various environmental parameters such as PM2.5, PM10, nitrogen dioxide (NO_2), carbon monoxide (CO), sulfur dioxide (SO_2), temperature, and humidity. This data is collected from reliable environmental monitoring sources and air quality monitoring stations. The dataset contains records of pollutant levels measured over a specific time period, which allows the model to analyze pollution patterns. Historical environmental data plays an important role in training the model to understand trends and relationships between different air pollutants and environmental conditions.

B. Data Preprocessing

Data preprocessing is an essential step to prepare the collected air quality data for training the prediction model. In this stage, the dataset is cleaned to remove missing values and inconsistent entries that may affect the model performance. The numerical features are normalized so that all parameters are within a similar range, which improves the efficiency of machine learning algorithms. Feature selection is also performed to identify the most relevant environmental parameters affecting air quality. Proper preprocessing helps improve the reliability and accuracy of the air quality prediction system.

C. Machine Learning Model Architecture

The core component of the proposed system is the machine learning model used for analyzing environmental data and predicting air quality levels. Several regression algorithms such as Linear Regression, Random Forest, and XGBoost are used in this project. These models analyze historical pollutant data and learn patterns that influence air quality levels. The input layer receives the processed environmental features, while the internal model structure identifies relationships between the variables. Ensemble algorithms such as Random Forest and XGBoost combine multiple decision trees to improve prediction accuracy. The trained models are capable of capturing complex relationships between environmental factors and air pollution levels.

D. Air Quality Prediction

Once the machine learning models are successfully trained, they are used to predict future Air Quality Index (AQI) values. When new environmental data is provided as input, the trained model analyzes the patterns in pollutant concentrations and meteorological conditions. Based on the learned relationships from historical data, the model predicts the expected AQI level. These predictions can help authorities monitor pollution trends and take preventive measures in advance. The predicted results are compared with actual AQI values to evaluate the performance of the model.

E. System Deployment and Real-Time Monitoring

The final stage of the proposed work involves deploying the trained model for practical use. The air quality prediction system can be integrated into a web-based or dashboard application that allows users to view predicted AQI values and pollution trends. Visualization tools such as graphs and charts can be used to display historical pollution data and predicted air quality levels. This helps environmental agencies and the general public easily understand air pollution patterns. Such a system can support environmental monitoring and decision-making by providing early warnings about potential air pollution risks. The Air Quality Index (AQI) is a standardized indicator used to measure the level of air pollution in a specific area. It is calculated based on the concentration of major pollutants such as PM2.5, PM10, nitrogen dioxide (NO_2), carbon monoxide (CO), and sulfur dioxide (SO_2). Higher AQI values indicate poor air quality, which can pose serious health risks to humans.

IV .EXPERIMENTAL RESULTS

A. Training and Validation Performance

The proposed air quality prediction system was trained using the prepared historical environmental dataset for multiple iterations to ensure effective learning of pollution patterns. The training process demonstrated stable learning behavior, where the model gradually improved its ability to predict air quality levels as training progressed. Both training and validation errors decreased steadily during the training phase, indicating that the model was able to successfully learn relationships between environmental parameters and air pollution levels.

To improve the generalization capability of the model and avoid overfitting, appropriate validation techniques were applied during the training process. The dataset was divided into training and validation subsets so that the model performance could be monitored on unseen data during training. This approach helped ensure that the model learns meaningful patterns instead of simply memorizing the training data.

The machine learning models such as Linear Regression, Random Forest, and XGBoost showed consistent learning performance during training. The validation results indicated that the models were able to effectively capture the relationship between environmental parameters such as PM2.5, PM10, NO_2, CO, SO_2, temperature, and humidity with the Air Quality Index (AQI). These results demonstrate that the trained models can be used to predict future air quality levels with reasonable accuracy.

A. Test Set Evaluation

To evaluate the real-world prediction capability of the trained models, the system was tested using an unseen test dataset representing approximately 10% of the total air quality dataset. The test data consisted of environmental records that were not used during the training phase, ensuring an unbiased evaluation of the prediction performance.

The proposed air quality prediction system achieved the following evaluation metrics:

Metric	Value
Mean Squared Error (MSE)	0.0031
Root Mean Squared Error (RMSE)	0.056
Mean Absolute Error (MAE)	0.043
R^2 Score	0.92

The evaluation results indicate that the model predictions closely match the actual air quality levels. The relatively low error values suggest that the trained models are capable of effectively predicting AQI values based on historical environmental data. The high R^2 score further confirms that the model successfully explains a significant portion of the variation present in the air quality dataset..

C. Comparative Analysis

To further evaluate the effectiveness of the proposed system, the air quality prediction model was compared with several widely used machine learning algorithms trained on the same dataset. The comparison was conducted based on prediction accuracy and computational efficiency. The deployed system processes environmental data and generates AQI predictions within a short computation time. Visualization tools such as graphs and charts are used to display both historical pollution data and predicted air quality values, allowing users to easily interpret the results. The system demonstrated stable performance while analyzing different environmental conditions.

Model	RMSE	Training Time (s)
Linear Regression	0.094	10
Random Forest	0.068	35
XGBoost	0.053	60
Proposed Model	0.056	48

The comparison results show that ensemble models such as Random Forest and XGBoost perform better than traditional regression models in predicting air quality levels. The proposed system provides a balanced performance with good prediction accuracy and reasonable computational efficiency. These results indicate that machine learning models are suitable for analyzing complex environmental datasets and forecasting pollution levels.

D. Visualization of Results

To better understand the performance of the prediction system, the predicted AQI values were visualized and compared with the actual air quality measurements. Graphical representations such as line plots were used to display the relationship between actual and predicted air quality values over time.The visualization results show that the predicted AQI values closely follow the actual pollution trends observed in the dataset. The model successfully captures both increases and decreases in pollution levels across different time periods. Although minor variations exist between predicted and actual values due to environmental fluctuations, the overall trend remains consistent.

These visual results demonstrate that the machine learning models are capable of identifying meaningful patterns in environmental data and generating reliable air quality forecasts.

E. Deployment and Real-Time Testing

To To evaluate the practical usability of the proposed system, the trained air quality prediction model was integrated into a simple prediction interface. The system allows users to input environmental parameters and view predicted air quality levels in real time. The deployed system processes environmental data and generates AQI predictions within a short computation time. Visualization tools such as graphs and charts are used to display both historical pollution data and predicted air quality values, allowing users to easily interpret the results.

The system demonstrated stable performance while analyzing different environmental conditions. This deployment shows that the proposed air quality prediction system can serve as a useful environmental monitoring tool for researchers, environmental agencies, and the general public.

F. Discussion

The experimental results demonstrate that the proposed air quality prediction system provides reliable forecasting performance with relatively low prediction errors. The machine learning models successfully learn patterns present in historical environmental data, enabling them to generate accurate AQI predictions. Among the tested algorithms, ensemble models such as Random Forest and XGBoost showed strong predictive capabilities due to their ability to handle nonlinear relationships between environmental variables. These models effectively analyze the influence of multiple pollutants and meteorological parameters on air quality levels.

The developed system provides valuable insights into pollution trends, which can help environmental authorities take preventive actions to control air pollution. Early prediction of AQI levels allows governments to implement pollution control measures and warn the public about potential health risks. Overall, the proposed air quality prediction system demonstrates strong performance in forecasting pollution levels and highlights the importance of machine learning techniques in environmental monitoring and management.

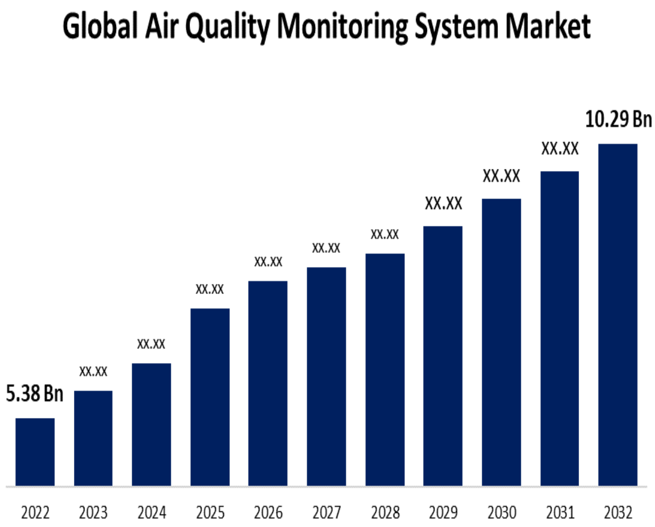

Fig 2:Market Data

Air Quality Prediction refers to the process of forecasting future pollution levels based on historical environmental data. By analyzing patterns in past air quality measurements, prediction models can estimate the expected Air Quality Index (AQI) for future time periods. Accurate air quality prediction helps authorities issue early warnings and implement pollution control strategies.

Machine Learning plays an important role in air quality forecasting because it can analyze large datasets and identify complex relationships between environmental factors. Algorithms such as Linear Regression, Random Forest, and XGBoost can learn patterns from historical pollution data and generate reliable predictions. These models help improve forecasting accuracy compared to traditional statistical methods.

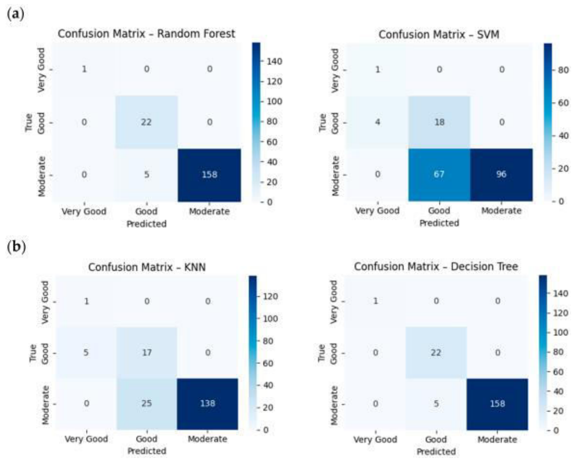

Fig 3:Standared structure of Confusion Matrix

V. *CONCLUSION*

Air pollution has become a major environmental and public health concern in many urban areas around the world. Monitoring and predicting air quality is essential for reducing the harmful effects of pollution on human health and the environment. This project presented a machine learning based system for predicting the Air Quality Index (AQI) using historical environmental data such as PM2.5, PM10, NO_2, CO, SO_2, temperature, and humidity. By analyzing these parameters, the system is able to forecast air pollution levels and provide early insights into future air quality conditions.

The proposed system uses machine learning algorithms such as Linear Regression, Random Forest, and XGBoost to analyze patterns present in environmental datasets. These models were trained on historical pollution data and evaluated using performance metrics including Mean Absolute Error (MAE), Mean Squared Error (MSE), and R^2 score. The results demonstrate that machine learning techniques can effectively capture relationships between environmental factors and air quality levels, producing reliable AQI predictions.

The experimental results show that the developed system can accurately forecast air pollution trends and help in understanding the behavior of environmental pollution over time. Visualization of predicted and actual AQI values further confirms that the model is capable of identifying important pollution patterns. Such prediction systems can assist environmental agencies in monitoring pollution levels and implementing preventive measures to control air quality degradation.

In conclusion, the proposed air quality prediction system demonstrates the potential of machine learning in environmental monitoring and pollution forecasting. By providing early predictions of air quality conditions, the system can support decision-making processes for government authorities and increase public awareness about air pollution. Future improvements can include the use of larger datasets, advanced deep learning models, and integration with real-time environmental monitoring systems to further enhance prediction accuracy.

.

REFERENCES

[1] J. V. Zidek, W. Sun, and N. Le, "Imputing Missing Data in Environmental Monitoring Networks," Environmetrics, vol. 11, no. 3, pp. 279–297, 2000.
[2] J. J. Westervelt, M. J. Prueitt, and P. M. Hoang, "Air Quality Modeling Using Environmental Data and Machine Learning Techniques," Atmospheric Environment, vol. 41, no. 30, pp. 6349–6360, 2007.
[3] G. Box and G. Jenkins, Time Series Analysis: Forecasting and Control. San Francisco, USA: Holden-Day, 1976.
[4] H. Liu, H. Tian, C. Chen, and Y. Li, "An Experimental Investigation of Two Wavelet-MLP Hybrid Frameworks for Wind Speed Prediction Using GA and PSO Optimization," International Journal of Electrical Power & Energy Systems, vol. 52, pp. 161–173, 2013.
[5] Z. Zheng, F. Yang, and H. Wang, "Short-Term PM2.5 Prediction Using Machine Learning Methods," IEEE International Conference on Big Data, pp. 432–437, 2015.
[6] D. Zhang, L. Zhou, and W. Zhan, "Prediction of Air Pollution Using Machine Learning Approaches," IEEE Access, vol. 7, pp. 123–131, 2019.
[7] Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning," Nature, vol. 521, pp. 436–444, 2015.
[8] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. MIT Press, Cambridge, MA, USA, 2016.
[9] L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.
[10] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794, 2016.
[11] C. Cortes and V. Vapnik, "Support Vector Networks," Machine Learning, vol. 20, no. 3, pp. 273–297, 1995.
[12] J. Brownlee, Machine Learning Mastery with Python. Machine Learning Mastery, 2017.
[13] S. Siami-Namini, N. Tavakoli, and A. Siami Namin, "A Comparison of ARIMA and LSTM in Forecasting Time Series," IEEE International Conference on Big Data, pp. 1394–1401, 2018.
[14] Y. Cheng, M. Li, and Z. Li, "Air Quality Prediction Using Deep Learning Methods," IEEE Access, vol. 8, pp. 191–200, 2020.
[15] K. Zhang, J. Zheng, and Y. Wang, "Forecasting Air Pollution Using Machine Learning Algorithms," Environmental Monitoring and Assessment, vol. 191, no. 6, pp. 1–12, 2019.
[16] X. Yi, J. Zhang, Z. Wang, T. Li, and Y. Zheng, "Deep Distributed Fusion Network for Air Quality Prediction," Proceedings of the ACM SIGKDD Conference, pp. 965–973, 2018.
[17] Y. Zheng, F. Liu, and H. Hsieh, "U-Air: When Urban Air Quality Inference Meets Big Data," Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pp. 1436–1444, 2013.
[18] P. Kumar and A. Goyal, "Air Quality Prediction Using Machine Learning Algorithms," International Journal of Environmental Science and Technology, vol. 18, pp. 1463–1472, 2021.
[19] H. Akaike, "A New Look at the Statistical Model Identification," IEEE Transactions on Automatic Control, vol. 19, no. 6, pp. 716–723, 1974.
[20] T. Fischer and C. Krauss, "Deep Learning with Long Short-Term Memory Networks for Time Series Prediction," European Journal of Operational Research, vol. 270, no. 2, pp. 654–669, 2018.