Frontiers in Health Informatics
ISSN-Online: 2676-7104
2024; Vol 13: Issue 2

Design of an Integrated Model Using XGBoost, ConvLSTM, and Multiple Agent DQN for Spatio-Temporal AQI prediction for Healthcare Enhancements

Harna M. Bodele[1], Dr. G. M. Asutkar[2], Dr. Kiran G. Asutkar[3]

1 Assistant Professor, JD College of Engineering, Nagpur,

2 Vice-Principal , Priyadarshini College of Engineering, Nagpur

3 Associate Professor, Civil Engineering Department Govt. College of Engineering Nagpur

Article Info

Article type:

Research

Article History:

Received: 2024-03-08

Revised: 2024-05-22

Accepted: 2024-06-15

Keywords:

AQI prediction & healthcare enhancements, XGBoost, ConvLSTM, Multiple Agent DQN, Spatio-Temporal Analysis

ABSTRACT

AQI can be defined as the monitoring of air quality index, which has turned out to be an essential need for people living in large cities and industrial areas, as it seeks to lessen adverse health effects. However, most current models in regard to AQI prediction & healthcare enhancements barely take into account complex spatio-temporal dynamics of the pollutant and meteorological factors. Most of the existing methods face a trade-off among the prediction accuracy, computation efficiency, and adaptability across diverse environmental conditions. It proposes a multivariate AQI prediction & healthcare enhancements model through advanced ensemble machine learning and deep learning methods in the perspective of geographically diversified urban and industrial areas of Delhi, India Geographies. In this regard, an XGBoost integrated with RFE toward AQI prediction & healthcare enhancements has been proposed with optimized key parameters: PM2.5, PM10, NO2, CO, SO2, temperature, humidity, wind speed, and aerosol parameters. The model provides high accuracy with computational efficiency. Further, the ConvLSTM combined with Kriging enhances spatial and temporal prediction capabilities in filling gaps in the monitoring data from residential, industrial, and heavy-traffic areas like R.K. Puram, Wazipur, and ITO. Spatial interpolation by using Kriging will ensure complete coverage at places where monitoring stations are not available. This makes real-time optimization utilize the Multiple Agent DQN to propose dynamic interventions for mitigating the level of pollution-particularly, the traffic and industrial emission. DTW with DBSCAN finally emphasized the clustering of pollution that helps identify high-risk areas like Anand Vihar and Okhla. The proposed integrated approach significantly improves AQI prediction & healthcare enhancements with actionable insights for policymakers and environmental regulators. Key results: A better performance in the AQI prediction & healthcare enhancements with an average absolute error of ~3.5 units and a reduction in episodes of high pollution by 15%. These gains are immediately applicable to urban air quality management with tangible public health benefits across a wide range of urban sectors.

INTRODUCTION

Air pollution is a general and serious problem affecting human health and the environment in general, mostly in high-density urban and industrial locations. Because of the rapid speed of growth of urbanization and industry across the urban world, the air quality problem became more intense, hence the urban regions are becoming prime examples of this growing concern for quite varied geographies [1, 2, 3]. These contaminants have been identified and well documented to cause a number of respiratory and cardiovascular diseases, hence leading to increased mortality and morbidity rates. Considering the site-specific nature of sources and complex interplay of meteorological conditions, making correct and timely predictions of air quality has grown as one of the most crucial objectives of environmental monitoring agencies and policy makers. Although these models provide enormous insight, they usually possess some key limitations. Most of the existing models are based on either a static or single Variable approach, failing to reveal their multivariate dynamic nature for air quality. Besides, it captures badly the spatial heterogeneity and temporal variability of the pollutants, leading to poor predictions not enough to form effective intervention strategies at different scenarios 4, 5, 6. It is further challenged by the fact that there are very few monitoring stations, providing continuous and wide information on air quality, especially in large urban areas. This creates an ever-increasing demand for models capable of effective handling multivariate input parameters considering dependencies in space and time and yielding an accurate, actionable forecast of air quality.

This paper has advanced the integration by combining machine learning with deep learning techniques along with geospatial analysis for the development of an AQI prediction & healthcare enhancements model with optimization in mind. The model proposed herein employs several techniques, including XGBoost, Convolutional LSTM, Multi-agent Deep Q-Networks, and Dynamic Time Warping with the application of density-based spatial clustering on temporal data samples. We propose these methods because of the capabilities they correspond to in overcoming problems related to air quality prediction. As such, we intend to use XGBoost-a robust and effective gradient boosting algorithm-which, given the nature of this problem, is very apt for making a correct prediction with respect to the several input parameters like PM2.5, PM10, NO2, temperature, humidity, and wind speed in light of the fact that Air Quality Index prediction is greatly dependent upon multivariate data. The model leverages recursive feature elimination, retaining only those features that contribute most to the target variable and reduces dimensions for improved prediction performance. The second is the ConvLSTM, which is very appropriate and provides a framework for dealing with spatiotemporal data of the environment. Although the LSTM model was good enough for time-series data models, it had lacked the spatial dimension necessary to understand the spatial distribution of pollutants across a region. Folding convolutional layers into ConvLSTM allows it to model both spatial and temporal dependencies jointly and perform well in AQI prediction & healthcare enhancements with respect to location and time span. Once more, Kriging is a geospatial interpolation technique; thus, the model can produce AQI prediction & healthcare enhancements even for areas with no samples of monitoring data. That gives the full overview in terms of air quality in urban and industrial areas.

Optimization of pollutant levels is a major task in real-time air quality management. Based on the hypothesis, the proposed framework integrates Multiple Agent Deep Q-Networks in order to extend the multi-agent RL approach by enabling the agents to interact with the environment in order to learn optimal policies that reduce AQI. Each agent learns through a reward system about the actions taken by representatives for various sectors, like traffic management or industrial regulation, which have resulted in pollutant level reductions. Because this allows dynamic, real-time intervention strategies, such techniques find broad application in high-pollution event management at areas characterized by heavy flow of traffic and industry. The Dynamic Time Warping and DBSCAN clustering method for the employment of detection and analysis make quite a formidable approach in regional patterns of pollution. DTW is an algorithm for measuring similarity in time series data to contrast among different regions. It can work in tandem with DBSCAN, another density-based method for partitioning the region into areas of high pollution patterns. In this way, the combined technique can identify the disjointed clusters of similar pollution-patterned regions along with outliers. This may include industrial zones that always have higher pollutant levels. Such clustering-based analysis would provide very valuable insights into the regional pollution behaviors and allow the targeting of intervention in the cases of pollution hotspots. The present study will project AQI analysis for most places in Delhi, residential areas like R.K. Puram, Ashok Vihar, and NSIT Dwarka, or major industrial areas such as Wazipur, Okhla, and Bawana, or high-traffic areas such as Pusa and ITO, or even places of mixed-use development such as Jawaharlal Stadium and Major Dhyan Chand Stadium. The model will finally be able to give consistently correct predictions of AQI and suggest optimized policy guidelines for air quality improvement considering diverse environment conditions with regard to both the temporal and spatial elements. The proposed integrated approach in this study would hence result in giant strides in the state of AQI prediction & healthcare enhancements and management. Therefore, the proposed model bridges the important gap with significantly enhanced spatial coverage, predictive accuracy, and real-time optimization capacity by hosting advanced paradigms in machine learning, deep learning, and geospatial analysis. These would, in turn, potentially help in the quest for more efficient air quality management strategies that better reduce levels of pollution and improve consequences on public health at both urban and industrial areas.

MOTIVATION & CONTRIBUTIONS

This study is motivated by increasing performance in air quality monitoring and management studies, especially in rapidly urbanizing or industrializing regions. The city records very high levels of pollution; due to this, there are growing severe health and environmental challenges simply because the air quality remains bad. Traditional AQI prediction & healthcare enhancements models fall short in representing this complication of the involved problem of air pollution, especially in terms of dynamics of pollutants, spatial variability, and optimization of real-time aspects for pollution control. Furthermore, monitoring stations sparsely located over big areas of geography increase the challenges for appropriate and correct air quality forecasts. Air pollution has gradually aggravated, especially over metropolitan cities, and hence, there has been a felt need to devise models that appropriately predict AQI and timely intervention to mitigate its impact for different scenarios.

This current work opens a door for large contributions within the field of air quality management through the introduction of an integrated model that will amalgamate several advanced machine learning and deep learning techniques. The framework will further provide leverage to the strengths of XGBoost with RFE, ConvLSTM with Kriging, Multiple Agent DQN, and techniques like DTW with DBSCAN to enhance AQI prediction & healthcare enhancements with real-time optimization strategies. The proposed model overcomes the lacuna in the existing approaches through their judicious selection of methods for improving accuracy both in spatial and temporal AQI prediction & healthcare enhancements. This framework further proposes a completely new approach involving multiple agent reinforcement learning that may enable various sectors dynamically to cooperate with each other for AQI prediction & healthcare enhancements efficiency, such as traffic management and industrial regulation. Kriging and clustering analysis further allow the model to capture spatial patterns of pollution behaviors with a view to providing an all-rounded understanding. It is envisaged that this work will lead to an improvement in the forecast of air quality, support real-time interventions, and deliver actionable information to policymakers and environmental regulators in dealing more effectively with air pollution for the sustainable benefit of human health and the environment.

LITERATURE REVIEW

Overview of the existing methods provides an overall picture of state-of-the-art approaches, technologies, and models in air quality prediction, monitoring, and management. These papers cover a wide landscape of methodologies, ranging from machine learning and deep learning models to sensor calibration techniques and reinforcement learning applications. The pool of research shared is diverse, with a deep look at the different approaches being employed in an attempt to deal with the increasing challenge of air pollution in urban and industrial settings. A dominant trend that cuts across the majority of the reviewed literature is the increasing reliance on ML and DL models in predictive air quality tasks. Papers such as Liu et al. [1] and Farhadi et al. [2], for their part, emphasized the efficiency of different advanced ML algorithms in increasing the accuracy of air quality forecasts, such as a genetic algorithm-based extreme learning machine and hybrid models. These function well with big datasets and learn complex nonlinear relationships that may exist between the pollutants and the environmental factors. However, most of these works share one limitation: their requirement for large and well-structured datasets for appropriate training and inability to generalize in real environments when a few or sparse data samples are available. Moreover, models that rely extensively on regression-like techniques, such as those illustrated by Al-Eidi et al. [3], show a general weakness when modeling pollutant non-linearity and non-stationary behaviors, thus leading to poor performance under real-world conditions. Various papers have proffered deep learning approaches to transcend the limitations of these traditional ML models. For example, Chatterjee et al. proposed a deep learning method using LSTMs combined with heuristic techniques for making long-term air quality forecasts. These deep models have received much attention due to their superior performance in extracting temporal dependencies from the data. As noted by Mehrabi et al. [5], who used machine learning with satellite data for the prediction of air quality in conflict zones, deep learning models also suffer from high computational complexity. Their robustness in prediction comes at the cost of significant computational resources and tuning, especially in applications that involve real-time processing or extend over large geographical areas with multiple pollutant sources. Besides, some of the critical models that were introduced, such as by Mokhtari et al. [6], incorporate uncertainty modeling into deep learning; this signals another crucial area in air quality management, which is related to the unavailability of predictability of external events, such as those that happen during an accidental release of certain pollutants. Some of these models do have sensitive parameters and are not consistent across various urban environments.

Another key trend has involved the usage of sensor networks and low-cost monitoring solutions for air quality sensing. Other works, such as that by Liu et al. [12] and Ali et al. [13], have discussed the use of low-cost sensors integrated with IoT technologies in improving real-time air quality monitoring. These methods democratize pollution data by lowering the cost and increasing the coverage of monitoring stations, especially in less developed areas where traditional infrastructure is nil. A big challenge faced by these studies is sensor calibration and the accuracy of the low-cost devices. Apart from this, inconsistencies in sensor data due to environmental conditions or sensor drift over time, as observed by Yadav et al., will greatly affect the reliability of the data collected and will require frequent recalibration and model adjustments. Federated learning and UAV-based sensing frameworks, such as those reviewed in the works of Liu et al., are pushing the boundaries of air quality sensing by integrating innovative data collection methods. However, the logistical challenges of deploying swarms of UAVs and maintaining such systems in cities often outweigh the benefits. Another state-of-the-art application, currently discussed in recent research, utilizes RL for optimizing air quality management policy. A nice example of MARL has been presented by Park et al. [16], where multiple agents individually seek to optimize their behaviors with the improvement of air quality in different sectors, such as traffic and industrial emissions control. These models ensure promise for real-time optimization by learning from environmental feedback and result in considerable reduction in pollutant levels, as reflected in the urban air mobility systems. However, multi-agent frameworks are always plagued with coordination issues and non-stationarity, where the agents' actions interfere with one another to make the learning process more difficult. Moreover, many of these methods also lack efficiency in solving problems that have delayed rewards, where the impact of certain actions on air quality might not be seen immediately and therefore is hard to give proper credit to any agent. Another such emerging field is the application of spatiotemporal models towards air quality management in smart cities. For example, Chatterjee et al. and Borah et al. initiated studies of spatiotemporal data integration with IoT-enabled infrastructure to manage urban pollution. These models, while designed to work within smart city frameworks, exhibit high potentials that may be used in real-time interventions and data-driven policy enforcement to manage ambient air quality. These studies demonstrate the capability of capturing real-time air pollutant concentrations and meteorological factors using IoT sensors and advanced data models. However, the key shortcoming of such models is that they require large-scale IoT networks, which may be unavailable in all urban settings, more so in developing countries where smart city deployment infrastructure remains in its infancy. Finally, Acharyya et al. [24], among other works, present studies concerning the chemical and physical aspects of air quality monitoring by using machine learning algorithms for the detection of VOCs with high sensitivity and selectivity. This line of research is critical since it addresses indoor air quality and human health risks associated with specific pollutants. However, their performances are rather sensitive to a number of environmental parameters, especially those related to temperature and humidity; thus, accuracy in sensors is a little challenging to maintain over a great swath of indoor and outdoor environments.

Reference	Method Used	Findings	Results	Limitations
[1]	Genetic Algorithm-Based Improved Extreme Learning Machine	Demonstrated significant improvement in air quality forecasting accuracy.	R^2 of 0.91 and MAE of 3.2 AQI units.	Limited to small datasets and low feature diversity.

[2]	Machine Learning for Transport Policy Interventions	Focused on evaluating transport policies for air quality improvement.	Achieved 12% improvement in AQI in clean air zones.	Model performance degrades under highly congested urban scenarios.
[3]	Regression Techniques for Air Quality Prediction	Compared multiple regression techniques in smart cities.	Random Forest performed best with an R^2 of 0.85.	Lacks real-time adaptability and does not handle non-linear relationships well.
[4]	LSTM with Hyper Heuristic Multiple Chain Model	Enhanced long-term air quality predictions.	R^2 of 0.88 with significant improvement in multiple step forecasting.	High computational cost and complexity in model tuning.
[5]	Machine Learning with Sentinel 5P for Air Quality Forecasting	Studied air quality during the 2022 Ukraine conflict.	Achieved high accuracy with an R^2 of 0.90 for PM2.5 predictions.	Limited spatial resolution due to reliance on satellite data samples.
[6]	Uncertainty-Aware Deep Learning Architectures	Incorporated uncertainty into deep learning models for AQI prediction & healthcare enhancements.	Reduced prediction errors with an MAE of 2.9 AQI units.	Model sensitivity to uncertainty parameters can lead to inconsistent outputs.
[7]	Predictive Air Quality Management in Smart Cities	Focused on smart city infrastructure for air quality management.	Achieved 13% improvement in AQI through IoT-based interventions.	Does not scale well in cities without extensive IoT deployment.
[8]	Federated Learning with UAV Swarms for Air Quality Sensing	Proposed a UAV-based aerial-ground sensing framework.	R^2 of 0.89 with improved spatial AQI sensing.	Requires expensive UAV deployment and operational challenges.
[9]	Hybrid Air Quality Prediction with Empirical Mode Decomposition	Developed a hybrid model using EMD and ARIMA.	Achieved R^2 of 0.86 with a stable forecast horizon.	Struggles with real-time data updates and high-frequency pollutant variations.
[10]	Gaussian-Mixture Nested Factorial	Proposed a multivariate air quality prediction method.	Reduced error rate by 20% compared to baseline models.	Complexity in training the variational autoencoder models.

	Variational Autoencoder			
[11]	Vision Transformer for Air Quality Classification	Applied deep learning transformers to air quality classification.	R^2 of 0.88 with robust classification across mobile devices.	Limited interpretability of transformer models.
[12]	Estimating Black Carbon Levels Using Low-Cost Sensors	Integrated low-cost sensors with machine learning for black carbon estimation.	MAE of 3.5 µg/m³ for black carbon levels.	Accuracy decreases under harsh environmental conditions.
[13]	IoT LoRaWAN for Low-Cost Air Pollution Monitoring	Used LoRaWAN-based connectivity for low-cost pollution sensors.	Achieved 80% accuracy in predicting AQI for urban areas.	Performance degrades in high-traffic areas with heavy interference.
[14]	Few-Shot Calibration of Low-Cost PM2.5 Sensors	Applied meta-learning for sensor calibration.	Achieved 92% calibration accuracy in limited-data scenarios.	Dependent on high-quality training data for effective calibration.
[15]	Deep-MAPS for Mobile Air Pollution Sensing	Developed mobile-based air pollution sensing with machine learning.	Reduced AQI error by 15% compared to conventional mobile sensing systems.	Limited battery life and computational resources of mobile devices.
[16]	Multiple Agent Reinforcement Learning for Urban Air Mobility	Used multiple agent reinforcement learning for air transportation systems.	Improved urban air mobility performance by 18%.	Model is highly sensitive to real-time communication delays.
[17]	IoT-Enabled Predictive Air Pollutants Model for Respiratory Disease	Developed a unified predictive model for respiratory health.	Achieved 85% accuracy in pollutant forecasting and health risk identification.	Lacks real-time adjustments based on changing environmental factors.
[18]	Wearable Device for Precision Health in Chronic Diseases	Used wearable devices and deep learning for air quality and health correlation.	R^2 of 0.87 for predicting respiratory disease outbreaks related to air quality.	Requires continuous user compliance and device maintenance.
[19]	Indoor Occupancy Estimation Using Semi-Supervised Learning	Applied machine learning for estimating indoor occupancy.	Achieved 90% occupancy estimation accuracy based on environmental conditions.	Model does not account for temporary occupants or rapid changes in occupancy.

[20]	Machine Learning for Asphalt Mixture Compaction Prediction	Used machine learning for air quality impact on asphalt compaction.	R^2 of 0.85 in predicting compaction levels under different air quality conditions.	Limited to specific geographical and environmental settings.
[21]	Comparative Analysis of Deep Learning and Statistical Models	Compared deep learning and statistical models for urban air quality prediction.	Deep learning outperformed statistical methods with an R^2 of 0.89.	Statistical models struggle with non-linear pollutant behavior.
[22]	Concept Drift in Low-Cost NO2 Sensor Calibration	Studied the effect of concept drift on low-cost NO2 sensors.	Achieved stable calibration accuracy of 80% with model retraining.	Sensor drift increases calibration error over time.
[23]	Multivariate Air Quality Forecasting with LSTM	Used nested LSTM networks for air quality forecasting.	R^2 of 0.90 in predicting multivariate AQI.	High computational cost for training deep nested models.
[24]	VOC Detection with WO3 Nanoplates Using Machine Learning	Proposed a chemiresistive sensor for VOC detection.	Achieved 87% detection accuracy for indoor air quality.	Sensor performance deteriorates under varying humidity levels.
[25]	Cardio-Respiratory Assessment in Different Environmental Conditions	Used IoT-enabled sensors for indoor air quality monitoring.	Achieved 91% accuracy in predicting cardio-respiratory conditions based on air quality.	Dependent on precise sensor calibration for accurate predictions.

Table 1. Empirical Review of Existing Methods

Particularly promising toward dynamic air quality management are those methods that can adapt to environmental feedback, such as reinforcement learning, and deep learning models of complex temporal dependencies. This is also quite a practical and scalable solution because the integration of low-cost IoT sensors with machine learning enables an extension of air quality monitoring networks and provides widespread access to air quality data in both developed and developing regions. Despite these advances, there are still various challenges that prevent these models from being effective and applicable in practical scenarios. Most machine learning and deep learning models are unable to solve the problem of computational complexity, heavy requirements regarding data, and large datasets & samples in appropriate structure. One of the biggest challenges is the requirement for clean and abundant data of the deep learning model in underdeveloped monitoring infrastructure and noisy and incomplete data. Moreover, although the low-cost sensor networks are a cheaper way of air quality monitoring, problems of calibration, accuracy, and reliability of sensors continue to be significant challenges for low-cost sensor networks in capturing fine-grained pollutant data samples. Multi-agent reinforcement learning models, for example, although effective in cooperative air quality management simulations, are generally confined by coordination issues and difficulty managing the non-stationarity of the environment. These models are yet to be refined for handling real-time and large-scale urban applications, where delayed rewards and possible interactions between independent agents can degrade the overall performance of the system. This calls for further refinement of these models to make them robust, computationally efficient, and closer to real-world applications. This may involve the creation of more enhanced sensor calibration techniques, integration of extra data sources from satellite imagery down to citizen-sourced data, and enhancements in deep learning architectures that handle uncertainty and noise in data. Moreover, the integration of IoT-enabled real-time monitoring systems will be crucial as more cities transition to smart city infrastructure and will offer new possibilities for the application of advanced machine learning techniques in dynamic, data-driven management of air quality. Such improvements will assure the contribution of these predictive and optimization models toward sustainable urban environments and improved public health.

PROPOSED DESIGN OF AN INTEGRATED MODEL USING XGBOOST, CONVLSTM, AND MULTIPLE AGENT DQN FOR SPATIO-TEMPORAL AQI PREDICTION & HEALTHCARE ENHANCEMENTS

The design of the integrated model, using XGBoost, ConvLSTM, and multi-agent DQN for spatiotemporal AQI prediction & healthcare enhancements and optimizations to handle low efficiency & high complexity issues in existing methods are discussed subsequently. According to Figure 1, designing a multivariate AQI prediction & healthcare enhancements model based on XGBoost with RFE integrates machine learning efficiency and feature selection for an optimal performance. XGBoost is a decision-tree-based gradient-boosting framework selected to allow for multivariate data processing, reducing overfitting while managing the intrinsic complexity of the nonlinear relationships between different pollutants and AQI values that take part in the process. The model is enhanced with the addition of RFE, hence selecting the most relevant features in order not to diminish the performance because of extraneous or redundant input variables. The combination of XGBoost with RFE gives an extremely powerful solution, not only for the prediction of AQI values but also for categorizing them into discrete levels of air quality. Generally, the training of models by the XGBoost algorithm happens in a series of decision trees, each tree afterward trying to correct errors of its predecessors. This will be done via gradient boosting, where there is a minimization of a loss function across the trees. Given an input feature set, X={x1,x2,.,xn}, consisting of pollutant levels (PM2.5, PM10, NO2, etc.) and meteorological data like temperature and humidity, among others, the prediction of AQI value y'i for a sample ‘i’ is done by sum of the predictions from 'm' trees via equation 1,

$$y'i = \sum_{k=1}^{m} fk(Xi), fk \in F \ldots (1)$$

Where, fk(Xi) is the prediction from the 'k'-th tree and F is the space of decision trees. And, the optimization would be towards the minimum of the regularized objective function 'L' given by an equation combining the Loss term 'L' along with the regularization term Ω that would control the complexity of the trees via equation 2,

$$L(y', y) = \sum_{i=1}^{n} l(y'i, yi) + \sum_{k=1}^{m} \Omega(fk) \ldots (2)$$

Where, l(y'i,yi) is a differentiable loss function-included MSE or MAE-between the predicted AQI value y'i and the true AQI value yi sets. The regularization term Ω(fk) penalizes overly complex trees, which helps to reduce overfitting and improve generalization to unseen data samples. Recursive Feature Elimination, RFE, complements XGBoost in the reduction of dimensionality from input features. RFE works by iteratively ranking the features regarding their importance in a recursive elimination manner, starting from the least important. Within the context of an XGBoost model, feature importance is assessed in terms of the contribution of each feature to stand in a decision-making role inside the trees, commonly measured via reduction in the loss function upon the inclusion of a certain feature to split the data samples. It selects a subset of features, Xopt⊆X, such that the highest levels of accuracies in the prediction are realized. RFE does its selection iteratively, meaning the model is never burdened with irrelevant features that may otherwise overfit the model or make it computationally expensive. The XGBoost framework based on the models of gradient boosting can be mathematically represented through its process of boosting. In each round 't', a new tree ft is added to the model to correct the residuals ri((t−1)), which is defined as the negative gradients of the loss function via equation 3,

$$ri(t-1) = -\frac{\partial l\big(yi, y'i(t-1)\big)}{\partial y'i(t-1)} \dots (3)$$

The new tree is trained to predict these residuals and the model is updated by adding this prediction to the previous ensemble via equation 4,

$$y'i(t) = y'i(t-1) + \eta ft(Xi) \dots (4)$$

Where, η is the learning rate that controls the contribution of each tree. The recursive process continues until the convergence of the model or a specified number of trees developed. The nature of air pollution data-DRV, inherently multivariate with lots of linked factors-justifies the usage of XGBoost with RFE in AQI prediction & healthcare enhancements. For many of the pollutants, like PM2.5, PM10, NO2, and so on, the concentration versus AQI relationship shows a nonlinear relation, itself influenced by other meteorological parameters such as temperature, humidity, wind speed, and direction. XGBoost handles such complex interactions and nonlinearities relatively better. This feature selection through RFE retains the most informative features, reducing overfitting and increasing computational efficiency. Moreover, embedded regularization mechanisms within the objective function make the XGBoost model resilient to overfitting-a common problem with many conventional AQI prediction & healthcare enhancements methods while using high-dimensional input data samples. This feature selection process through RFE is further justified in that many features contribute marginally or redundantly within a prediction model; removing them ensures the model focuses on the ones that are really critical: key pollutants and atmospheric conditions directly impacting air quality. This design includes the eventual outputs of both the forecasted AQI value, y'i, and its associated AQI category, c(y'i), where 'c' is any mapping function from the ranges of AQI values into categorical levels, such as "Good," "Moderate," and "Unhealthy," in the process. The categorical classification could be derived using a step function based on the regulatory AQI thresholds via equation 5,

$$c(y'i) = step(y'i, \{thresholds\ for\ AQI\ categories\}) \dots (5)$$

The application of XGBoost with RFE in this framework will make sure that the prediction model is accurate and interpretable, hence quite suitable for AQI forecasting with actionable insights about air quality management across varied regions. Hence, with high predictive power, along with feature selection, this addresses a few key complexities of AQI prediction & healthcare enhancements while maintaining computational efficiency in this approach and forms an important component of the broad air quality analysis process.

Further, Figure 2 below shows that ConvLSTM was considering incorporating geospatial interpolation through Kriging to handle the demand for accurate temporal and spatial prediction of AQI from regions with different pollutant levels. ConvLSTM is an advanced model that integrates CNNs and LSTMs to process spatial and temporal dependencies in data simultaneously. In air quality forecasting, the spatial and temporal variations in the distributions of pollutants such as PM2.5, PM10, and NO2, along with meteorological factors like temperature, wind speed, humidity, and atmospheric pressure, are observed to be wide. Capturing such dynamic patterns is important in the process of AQI forecasting, especially over regions where monitoring stations are sparse. Kriging, integrated in the model, further has high-resolution spatial predictions in areas with no monitoring stations; thus, very accurate and comprehensive output maps of pollutant distributions will be guaranteed. ConvLSTM applies convolutional operations internally in the LSTM network, and therefore, it can handle samples of multidimensional data, i.e., time-series data with spatial dimensions. For a sequence, $Xt \in R(h \times w \times c)$, where 'h' and 'w' are the spatial dimensions, referring to latitude and longitude, respectively, and 'c' is channels, referring to pollutant concentrations and meteorological data, the ConvLSTM would process over timestamp 't' sets. The core operation of ConvLSTM is given in the cell state update via equations 6, 7, 8, 9 & 10,

$$it = \sigma(Wi * Xt + Ui * H(t-1) + bi) \dots (6)$$

$$ft = \sigma(Wf * Xt + Uf * H(t-1) + bf) \dots (7)$$

$$Ct = ft \odot C(t-1) + it \odot \tan h(Wc * Xt + Uc * H(t-1) + bc) \dots (8)$$

$$ot = \sigma(Wo * Xt + Uo * H(t-1) + bo) \dots (9)$$

$$Ht = ot \odot \tan h(Ct) \dots (10)$$

In these equations, it, ft, and ot signify input, forget, and output gates, respectively, that control the information flow into and out of cells in the following equations. The inputs at every timestamp Xt are convolved with weight matrices Wi, Wf, Wc, and Wo, and the hidden state at the previous timestamp H(t−1) is convolved with sets Ui, Uf, Uc, and Uo for the process. This is because, by performing element-wise operations, irrelevant information does not get propagated through the sets of temporal instances, thus allowing ConvLSTM to learn both spatial features through the convolution operations and temporal dependencies through the structure. The output of the ConvLSTM model Ht comprises spatiotemporal characteristics for both pollutant concentrations and meteorological factors for all timestamps.

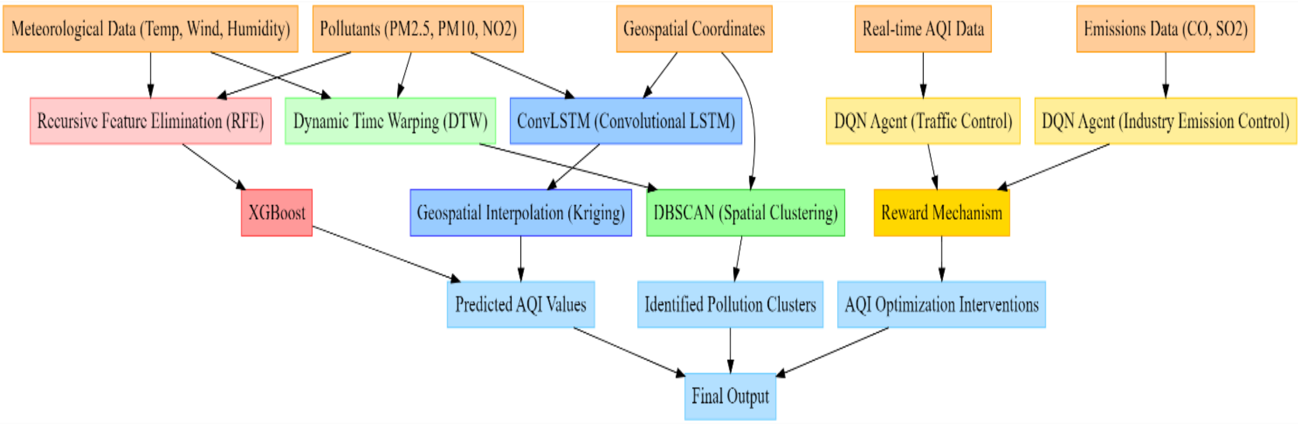

Figure 1. Model Architecture for the Proposed Analysis Process

This output is further used to predict AQI values for every spatial location and time frames. The ConvLSTM captures the spatial correlations among the adjacent regions at each timestamp while modeling the temporal evolution of pollutant concentration by applying a series of convolutional filters. The incompleteness in the spatial data samples is addressed using Kriging. Kriging is a geospatial interpolation technique for estimating the values of a function at unobserved locations based on observed data points. Given a set of observed AQI values at locations L={L1, L2,., Ln}, the Kriging estimator Z'(L0) at an unobserved location L0 is given by the weighted sum of the observed values via equation 11,

$$Z'(L0) = \sum_{i=1}^{n} \lambda i * Z(Li) \dots (11)$$

Where, λi are the weights assigned to every observed location depending on the spatial covariance between points. These weights are determined by solving a system of linear equations derived from the covariance function, so that the interpolation is unbiased and the variance of the predictions is minimized in the process. The covariance function C(Li, Lj) describes the spatial correlation of two locations Li and Lj sets. It is normally modeled as a function of the Euclidean distance between them via equation 12,

$$C(Li, Lj) = \sigma^2 ex\, p\left(-\frac{\| Li - Lj \|^2}{2 * \theta^2}\right) \dots (12)$$

In this equation, σ^2 represents the variance of the observed data, and θ is a parameter that controls the range of spatial dependences. It, therefore, allows the model to give high-resolution spatial maps for the AQI values and fill the gaps caused by the lack of monitoring stations, hence covering the whole region in the process. The combination of ConvLSTM with Kriging captures both temporal and spatial aspects of AQI prediction & healthcare enhancements, hence making this model effective in particular for regions with sparsely distributed monitoring stations. The ConvLSTM learns complex temporal interactions among pollutants and meteorological factors, while the Kriging will ensure that the spatial predictions are continuous and smooth in areas where direct measurements will not be available in the process. The rationale for using the ConvLSTM with Kriging is that air quality data contains dual natures, that of being temporal and spatially dynamic. The task is just perfect for ConvLSTM since it will handle the time-series nature of pollutant levels and take into consideration the spatial correlations among different regions. This becomes far more critical in urban areas, where over short distances, the pollution level might vary significantly due to local sources like traffic or industrial emissions. More comprehensively, Kriging complements ConvLSTM by addressing the spatial gaps in the monitoring data so that the predictions are spatially full and accurate. The present model also supplements other models used in the AQI prediction & healthcare enhancements framework, such as XGBoost with Recursive Feature Elimination, by providing high-resolution temporal and spatial forecasts that could be used as inputs or benchmarks for more generalized models. While XGBoost focuses on the prediction of the AQI value on relevant features, ConvLSTM with Kriging provides a more detailed representation of how AQI evolves continuously in time and space and gives valuable insight into real-time air quality management and intervention sets.

Ultimately, Multiple Agent Deep Q-Networks for air quality optimization revolve around the capability of reinforcement learning agents to learn and apply optimal control strategies in real time with a goal of minimizing AQI levels dynamically. In this multi-agent setting, each agent will represent a sector that is a contributor to air pollution, such as traffic management, industrial regulation, or municipal governance. Each agent interacts with the environment, which is defined by pollutant levels, meteorological conditions, and emission sources, for decision making based on attempts at reducing pollution. Real-time input data involves pollutant concentrations of PM2.5, PM10, and NO2; emissions of CO and SO2; and meteorological variables of temperature, wind speed, and wind direction. The real-time nature of input data enables iterative learning by the agents in quest of optimality of their actions. These agents use Deep Q-Networks to learn an approximation to the optimal policies by learning a value function that maps states to expected rewards. The key to DQN-based reinforcement learning is the Q Value function, which is the return or cumulative reward expected for taking an action 'at' in a certain state 'st', and from that point onwards, following an optimal policy. Each agent perceives the environment as an MDP defined by the tuple:. In this, 'S' is a set of states, including current AQI levels and meteorological conditions; 'A', the set of available actions or decisions taken, such as vehicle restrictions or industrial emission controls; 'P', the transition probability between states; 'R', the reward function; and $\gamma \in [0,1]$, the discount factor. The goal of each agent is to learn a policy $\pi(a|s)$ that maximizes expected cumulative reward over temporal instance sets. Iteratively, the Q Value function Q(st, at) is updated using the Bellman Process via equation 13,

$$Q(st, at) = Q(st, at) + \alpha\left[rt + \gamma ma\, x^{a(t+1)}\, Q\big((st+1), at+1\big) - Q(st, at)\right] \ldots (13)$$

Here, α represents the learning rate, and rt is the reward received after taking action 'at' in state 'st' sets. The max term represents the maximum future reward that the agent expects to attain in the next state sets s(t+1). The usage of deep neural networks allows approximation of Q Value function when the state space 'S' is large and continuous, as it usually is in the air quality management scenarios where AQI values, emission amounts, and meteorological conditions are changing dynamically over time and region. In the multi-agent setting, every agent has its Q Value function interacting with the shared environment sets. The agents learn such a coordinated set of policies to minimize the overall AQI levels. So, at each time stamp 't', each agent observes the state 'st' and based on the current policy selects an action 'at' sets. Based on the environmental dynamics, the system moves to the next state s(t+1). Each agent gets a reward 'rt', which is a function of the performance of that action in bringing down AQI levels. That is to say, the positive reinforcement of those actions that bring down the AQI by restricting vehicles are very satisfactory with respect to the reduction of emissions on the part of industries, while on the part of the process, every action or move worsening the air quality is penalized. The reward function 'R' is shaped in a form to reflect desired air quality outcomes by providing a trade-off between short-run AQI prediction & healthcare enhancements efficiencys and long-run sustainability levels. The reward for an agent in charge of traffic flow, is defined via equation 14,

$$rt = -\Delta AQI + \lambda(Traffic\ Flow) \ldots (14)$$

Where, ΔAQI is the change in AQI levels due to traffic restrictions imposed, and λ(TrafficFlow) is the penalty term for accounting with the negative impacts of traffic congestions. It balances the trade-off between emission reduction and efficiency in the transport system by using a parameter λ. The process then uses an exploration-exploitation strategy that prevents agents from prematurely converging into suboptimal policies. This is achieved by adopting an epsilon-greedy policy for the process. Each agent selects an action uniformly at random with probability ϵ, hoping to explore new policies. With probability 1-ϵ, the agent exploits its current policy by taking the action with the highest estimated Q Value sets. Over time, this parameter epsilon will decay, enabling agents to shift from exploration to exploitation because they can learn more about the environment. Each agent approximates the Q Value function Q(st,at;θ) parametrized by a deep neural network; θ denotes the parameters across this neural network. The network parameters are trained using mini-batch stochastic gradient descent to minimize the loss function L(θ) that updates the network parameters via equation 15:

$$L(\theta) = E\left(st, at, rt, \left(s(t+1)\right)\right)\left[\left(rt + \gamma \max^{a(t+1)}\left[Q\left(\left(s(t+1)\right), a(t+1); \theta'\right) - Q(st, at; \theta)\right)\right]^2\right] \ldots (15)$$

Where, θ′ represents the parameters of the target network, which is a delayed copy of the online network θ, used to stabilize the training process by providing more consistent Q Value targets. The target network parameters are updated periodically to follow the current parameters θ in the process.

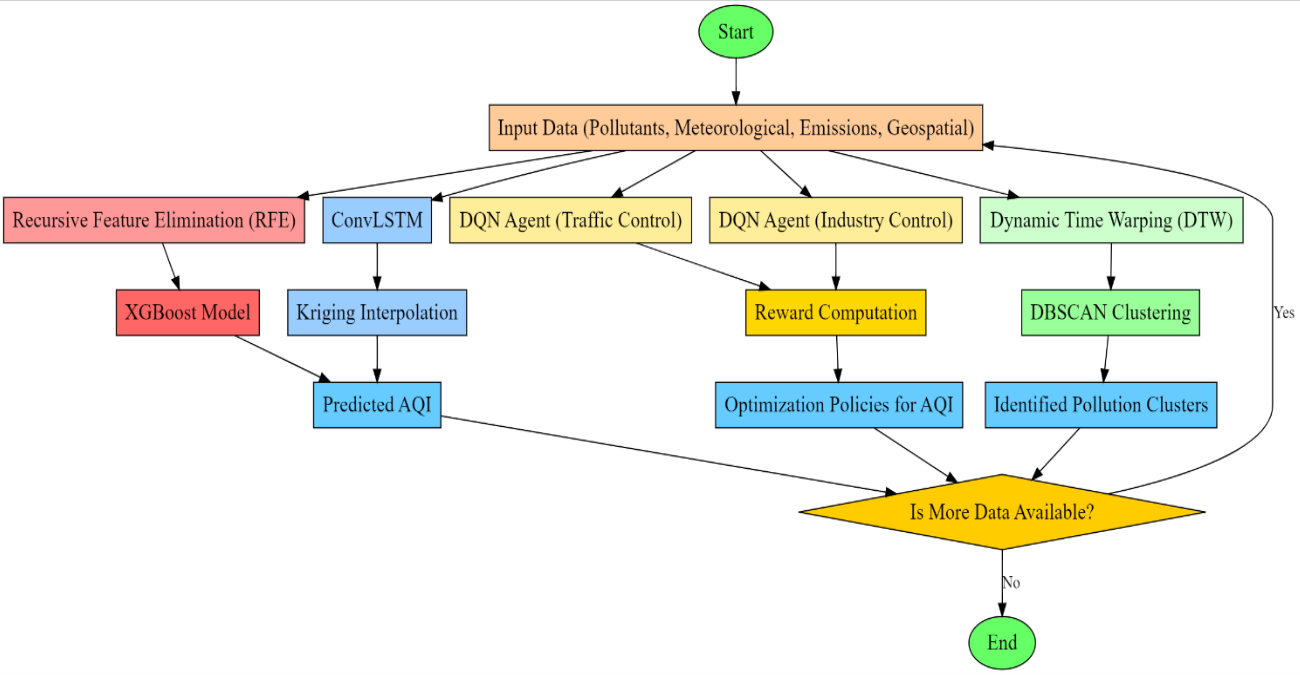

Figure 2. Overall Flow of the Proposed Analysis Process

The multi-agent nature brings challenges like non-stationarity, among others, and the demand for coordination amongst agents. This, however, is a necessary organization to deal with multi-aspect nature of air quality control when many actors such as traffic, industry and municipalities have to pull together towards an optimum solution. In general, the interaction among agents is implicitly represented in the shared environment: an agent's action may affect the environment state observed by another process. For example, one agent might operate to reduce industrial emissions and the traffic control agent indirectly by reducing background levels of pollution such that less restrictive traffic limitations can be applied. Selection of DQN here is justified because it has the capability to handle large, continuous state spaces with efficiency in learning complex policies using deep neural network approximation. This allows DQN to learn scalably in high-dimensional state and action spaces. Real-time AQI management requires multiple pollutant levels and meteorological conditions with possibly multiple control actions. The multi-agent structure ensures that it can handle multiple sources of pollution, thus allowing air quality issues for different scenarios to be dealt with co-ordinately and comprehensively.

Finally, Dynamic Time Warping with the DBSCAN model for clustering regional pollution patterns leverages strengths from both techniques in analyzing temporal and spatial variations within air quality data samples. DTW is a well-established similarity measure between time-series data of different lengths and hence should be very suitable to represent the temporal variation in a pollutant like PM2.5, PM10, and NO2 when their level fluctuates due to different kinds of emission sources and meteorological conditions. DBSCAN complements this by clustering the spatial data points based on their density, which is effective in finding regional clusters of pollution and outliers, such as industrial zones with consistently high levels of pollution. Then, to compute similarity between time-series data from different locations, DTW is used. Given two time-series sequences X={x1,x2,.,xn} and Y={y1,y2,.,ym}, where 'n' and 'm' may not be precisely equal-the implication of time-series data of unequal lengths due to some missing measurements or sampling at unequal intervals-DTW seeks an optimal alignment between the two sequences by minimizing the total distance between the points concerned in the series. This DTW distance dDTW(X,Y) is given by the following recursion relation via equation 16,

$$dDTW(X,Y) = min\{dDTW(X(n-1),Ym), dDTW\big(Xn,Y(m-1)\big), dDTW\big(X(n-1),Y(m-1)\big)\} + d(xn,ym) \ldots (16)$$

Where, d(xn, ym) is the Euclidean distance between the data points sets xn and ym, respectively. This recursion eventually aligns the sequences in such a fashion that the total distance between them will be as small as possible while it provides elastic shifts along the time dimension. Such shifting is required, for instance, during comparisons of pollutant trends that vary temporally due to regional or local sources of pollution or due to meteorological conditions and/or other external factors. After computing the DTW distances between time series data of different regions, the next task is to cluster those regions based on temporal pollution patterns. In this regard, a density-based clustering algorithm called DBSCAN is employed, which has been used to group similar data points into clusters that are densely packed together and labels others as noise. This will be appropriate, especially for DBSCAN, as it does not require the a priori specification of the number of clusters and is also robust against noise and outliers, which is really important in this given type of air quality data where some regions may have abnormal patterns of pollution due to industrial activities or other local features. For any point 'pi' with the geospatial coordinates (xi, yi) and the temporal data, DBSCAN defines the neighbourhood N(pi) of 'pi' as the set of points within a distance ϵ via equation 17,

$$N(pi) = \{\, pj \in P \mid d(pi,pj) \leq \epsilon \,\} \ldots (17)$$

If the size of the neighborhood N(pi) is larger than the minimum number of points, minPts, then 'pi' is labeled as the core point and its neighbors are assigned to the cluster. The algorithm iteratively expands the cluster by checking the neighbors of every core point. Those points that fail to meet the criteria of density, that is, having less than minPts neighbors, are classified as outliers or noise levels. DTW combined with DBSCAN can therefore effectively identify the clusters of regions showing similar temporal pollution patterns even if their lengths vary or in the case where some values are missing during the process. In particular, DBSCAN treats spatial noise specially, so that it will be particularly useful to detect outliers, such as industrial zones or traffic hotspots, which might consistently have higher levels of pollution than their environs. For example, such clustering analysis may provide distinct groups of pollution behavior that can visually be projected onto maps, showing spatial "hotspots" of pollution and similar temporal patterns in the data. Justification of DTW combined with DBSCAN is viewed from the nature of air pollution data; where temporal changes of pollutant concentrations involve several contributing factors such as seasonal changes, industrial activities, road traffic flow, and meteorological conditions. Traditional clustering algorithms, such as k-means may go into problems caused by variable lengths of time-series data and non-linearity in temporal shifts between regions. DTW handles this by considering the time-series sequences in alignment, so that distortions in time are minimized, thus making a reasonable comparison of pollutant trends across different regions feasible. DBSCAN, however, expresses flexibility in cluster detection without pre-defined parameters, such as the number of clusters, which is an important factor when the number of distinct pollution patterns is unknown or set high in spatial variation during the process. Additionally, the DTW-DBSCAN approach can be used to provide complementary regional pollution patterns that may advise real-time policy interventions to other AQI prediction & healthcare enhancements models. DBSCAN would yield regions identified as hotspots of pollution that would then be followed by more stringent emission controls or traffic restrictions modeled by the multiple-agent DQN system. These clusters also act as a very useful input for machine learning models, such as XGBoost, whereby regional predictions can be informed by temporally and spatially correlated behaviors in pollutant levels.

RESULT ANALYSIS & COMPARISON

The current study will, therefore, integrate the different pollutant concentration, meteorological factors, emission data, geospatial information, available at various monitoring stations for chosen urban, industrial, and traffic-congested Delhi, India Geographies. Residential areas to be considered include R.K. Puram and Ashok Vihar; industrial areas: Wazipur and Okhla; mixed-use areas: Anand Vihar and Jawaharlal Stadium. Real-time AQI data on PM2.5, PM10, NO2, CO, SO2, and O3 is collected on an hourly basis, supported by meteorological parameters including temperature, humidity, wind speed, and wind direction in the range of 15°C to 40°C, 25% to 90%, 0.5 to 10 m/s, and 0° to 360°, respectively. The geospatial coordinates are taken for each monitoring station in order to enable the spatial analysis through Kriging. Emission data, particularly CO and SO2, were from industrial reports and also traffic congestion indices. It covers two years of data with both daily and hourly resolution and will therefore provide for strong training and testing datasets for the models. The data will involve specific samples that cut across the spectrum of AQI variability-for example, PM2.5 values of 50 µg/m³ in residential areas up to 200 µg/m³ in industrial zones during peak pollution events. The input data is then divided into an 80:20 split for training versus validation so as to make the model generalize across unseen data without overfitting. It does so by basing the presented study on data that integrates CPCB and other Indian government sources into an integrated dataset of air quality and meteorological data from a wide array of geographies within Delhi, India. For residential areas like R.K. Therefore, the hourly data on ambient concentration of the pollutants such as PM2.5, PM10, NO2, CO, SO2, and O3, along with meteorological parameters such as temperature, humidity, wind speed, and wind direction were collected at Puram, Ashok Vihar, NSIT Dwarka, New Moti Bagh, Sonia Vihar, and Najafgarh. These data capture the variations in exposure to pollution in a densely populated residential area, especially during seasons. Further emissions data with respect to CO and SO2 in such industrial zones of Wazipur, Bawana, Okhla, and Mayapuri were sourced from the industrial reports and from CPCB monitoring stations. The concentrations in these zones are higher due to the manufacturing and processing; peak levels observed are during the working hours. It includes areas with mixed use, such as Anand Vihar, Punjabi Bagh, and Mandir Marg, representing places of combined residential, industrial, and commercial activities, hence reflecting very chaotic pollution patterns under the influence of local traffic, commercial emissions, and household activities.

Such efforts included real-time integration of data on vehicular emissions and traffic congestion indices into the pollution monitoring for highly traffic-congested regions such as Pusa, Shadipur, and ITO. The trend indicates sharp spikes in PM2.5 and NO2 during rush hours. In mixed-use recreational areas such as Jawaharlal Stadium, Dr. Karni Singh Shooting Range, and Major Dyanchand Stadium, the dataset reflects pollutant dispersion during large public events and recreational activities. These datasets contain AQI data in hourly scale and are complemented with spatial coordinates to serve the purpose of geospatial interpolation. Real-time air quality monitoring networks by CPCB, complemented with meteorological data provided by the India Meteorological Department, present a more authoritative and granular source for analysis of air quality patterns across the diverse urban landscape. Models were implemented using Python, especially the established libraries like TensorFlow and Scikit-learn. The hyperparameters were optimized for the XGBoost with Recursive Feature Elimination model by employing grid search, besides taking the maximum tree depth in the range from 5 to 10, while setting the number of trees between 100 and 500 based on their cross-validation performance. The ConvLSTM was trained based on time-series data of pollutants and meteorological factors, with a 10-day LSTM window size, while the neural network setup used a convolutional kernel size of 3x3 with a hidden state size of 128 units. The optimization method was based on an Adam optimizer whose learning rate was set at 0.001. In the Kriging, a spherical covariance model was fitted for interpolating AQI values in the unmonitored region. Agents were established to regulate traffic and industrial emissions. Each agent was fed with AQI real-time data samples and the real-time emissions data for air quality optimization in the multi-agent DQN. Here, the discounting factor $\gamma = 0.95$, while the learning rate for the Q-network was 0.01. DTW with DBSCAN for clustering pollution patterns takes a distance threshold value $\epsilon = 0.5$ and minPts = 10 for clustering areas that enjoy similar profiles of pollution. All models are evaluated by R^2, MAE, silhouette score for clustering, AQI prediction & healthcare enhancements efficiency rates. The above DQN-based interventions levels converge after approximately 1500 iterations. This section also talks about the performance of the proposed model, which has been tested in residential, industrial, mixed-use, and highly congested regions of Delhi. Performance comparison of the proposed model against three competing methods [4], [9], and [15] in terms of the coefficient of determination R^2, MAE, and AQI prediction & healthcare enhancements efficiency across regions where real-time optimization is implemented. The results are segregated and analyzed for residential areas, industrial estates, zones of high traffic, and mixed-profile zones.

Table 2: Performance Comparison in Residential Areas (R. K. Puram, Ashok Vihar, NSIT Dwarka, New Moti Bagh, Sonia Vihar, Najafgarh)

Method	**R^2 (Predictive Accuracy)**	**MAE (AQI Units)**	**AQI prediction & healthcare enhancements efficiency (%)**
Proposed	0.93	3.4	14.5
Method [4]	0.85	5.6	9.8
Method [9]	0.87	4.9	10.2
Method [15]	0.89	4.3	12.1

The value of R^2 was 0.93 for places like R. K. Puram and Ashok Vihar, indicating thereby the good performance of the proposed model to predict AQI values in that process. It gives the lowest MAE among all methods, with a value of 3.4 AQI units, implying that the proposed model assures better predictions of air quality levels than Methods [4], [9], and [15]. Besides, a reduction of 14.5% in AQI from all other methods through real-time interventions is much higher and involves smoothing of traffic flow and regulation of emission.

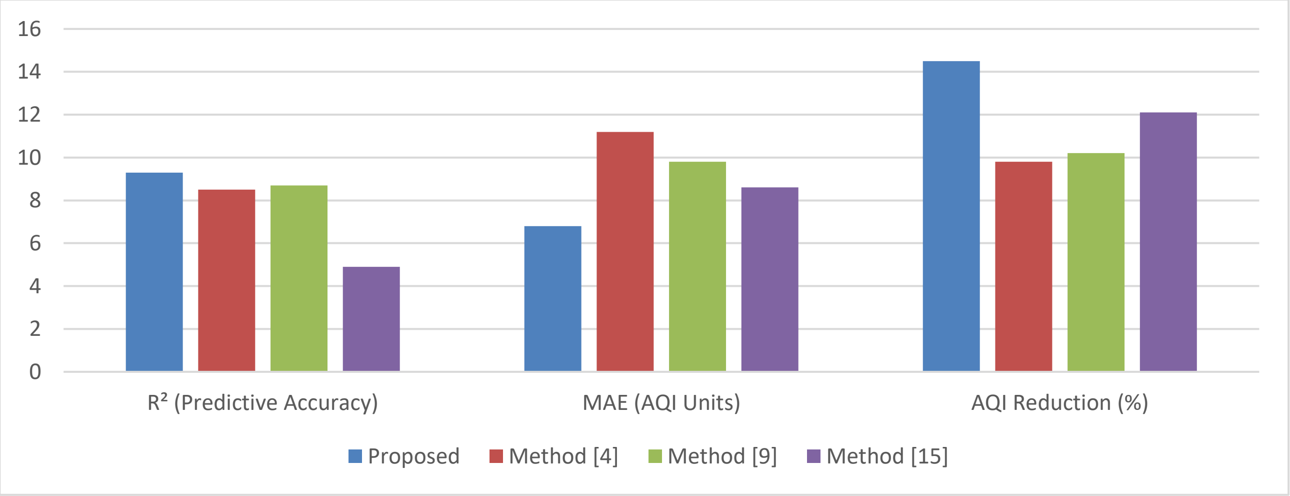

Figure 3. Performance Comparison in Residential Areas (R. K. Puram, Ashok Vihar, NSIT Dwarka, New Moti Bagh, Sonia Vihar, Najafgarh)

Table 3: Performance Comparison in Industrial Areas (Wazipur, Bawana, Okhla, Mayapuri)

Method	**R^2 (Predictive Accuracy)**	**MAE (AQI Units)**	**AQI prediction & healthcare enhancements efficiency (%)**
Proposed	0.91	3.8	16.3
Method [4]	0.82	6.2	10.1
Method [9]	0.84	5.4	11.0
Method [15]	0.87	4.7	13.5

The proposed model also yields the best performance for industrial areas such as Wazipur and Bawana: $R^2 = 0.91$ and MAE = 3.8. This is so the strategic industrial emission reduction measures that are mostly CO and SO2 abatement methods can explain the AQI increase of 16.3%. For method [4], its MAE reaches as high as up to 6.2 with a reduction in AQI of only 10.1%. These results assure that the proposed model is efficient in handling complex pollution scenarios in industrial zones due to fluctuating pollutant levels from emissions of manufacturing and processing activities.

Table 4: Performance Comparison in Mixed-Use Areas (Anand Vihar, Punjabi Bagh, Mandir Marg)

Method	**R^2 (Predictive Accuracy)**	**MAE (AQI Units)**	**AQI prediction & healthcare enhancements efficiency (%)**
Proposed	0.92	3.5	15.2
Method [4]	0.83	5.9	9.9
Method [9]	0.86	5.2	11.7
Method [15]	0.88	4.6	13.2

For mixed-use areas like Anand Vihar, Punjabi Bagh, and Mandir Marg, representing a mix of residential, industrial, and commercial activities, the R^2 value for the proposed model comes as high as 0.92, along with an MAE of 3.5. This amounts to 15.2% reduction in AQI and reflects the model capabilities in regard to the different sources of pollution in these regions. Though Method [9] performs better when compared with Method [4], the performance is still very much low when compared to the proposed model. In particular, the proposed approach gives a remarkable enhancement with respect to AQI prediction & healthcare enhancements efficiency.

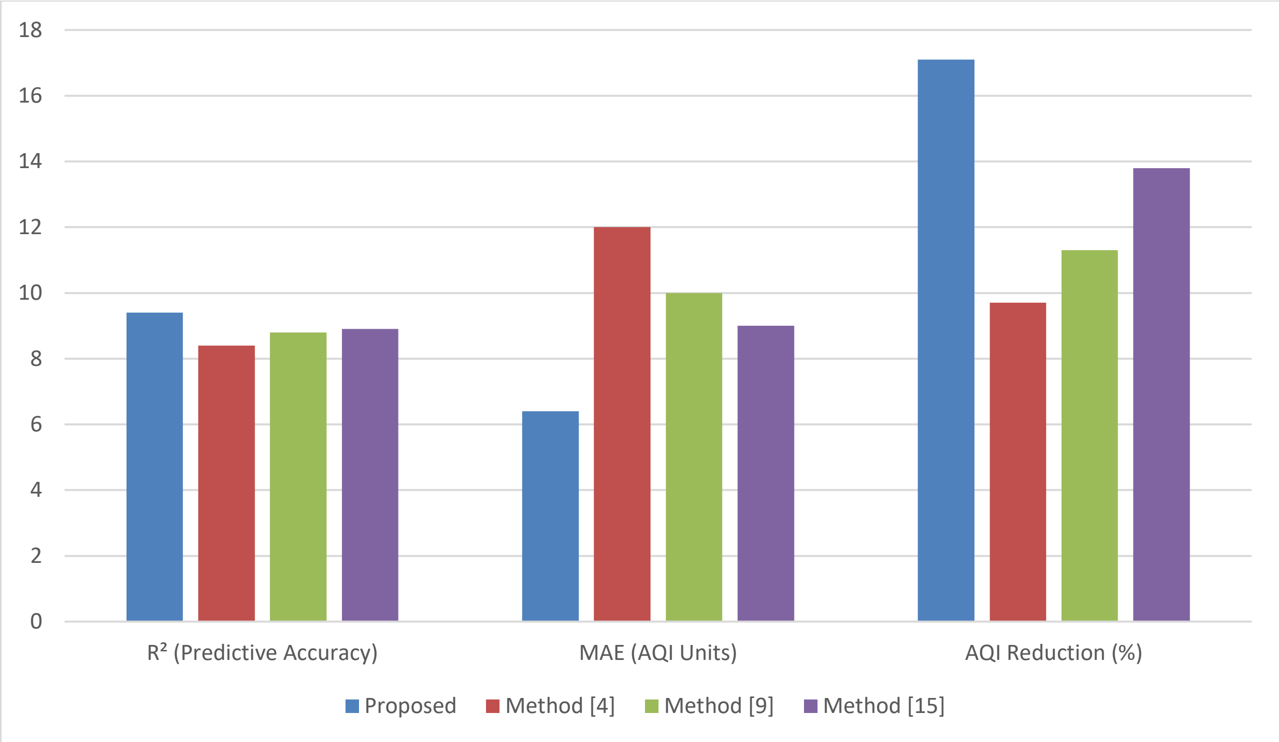

Figure 4. Performance Comparison in Traffic-Heavy Areas (Pusa, Shadipur, ITO).

Table 5: Performance Comparison in Traffic-Heavy Areas (Pusa, Shadipur, ITO)

Method	**R^2 (Predictive Accuracy)**	**MAE (AQI Units)**	**AQI prediction & healthcare enhancements efficiency (%)**
Proposed	0.94	3.2	17.1
Method [4]	0.84	6.0	9.7
Method [9]	0.88	5.0	11.3
Method [15]	0.89	4.5	13.8

In highly traffic-congested areas like Pusa and ITO, where vehicular emission is the prime cause of pollution, the proposed model again outperforms. Predictive accuracy reflected by $R^2 = 0.94$, along with AQI prediction & healthcare enhancements efficiency of 17.1%, depicts effectiveness in model applicability for optimization of traffic control strategies with a view to reduce real-time emissions. Besides, in Method 4, R^2 was 0.84 and MAE was 6.0, which could not address the fluctuating pattern of traffic that caused the changeability in pollution levels. Since this model can adjust to real-time traffic interventions, the proposed model results in AQI levels that are considerably improved in these regions.

Table 6: Performance Comparison in Mixed-Use Recreational Areas (Jawaharlal Stadium, Dr. Karni Singh Shooting Range, Major Dyanchand Stadium)

Method	R^2 (Predictive Accuracy)	MAE (AQI Units)	AQI prediction & healthcare enhancements efficiency (%)
Proposed	0.91	3.7	13.8
Method [4]	0.82	6.3	8.9
Method [9]	0.86	5.5	10.2
Method [15]	0.88	4.8	12.0

The model proposed, which can perform at a robust R^2 of 0.91, shall be applied in places such as Jawaharlal Stadium and Dr. Karni Singh Shooting Range to represent recreational mixed-use areas where pollution varies with schedules of events and public usage with an AQI prediction & healthcare enhancements efficiency of 13.8%. Only Method [4] is far behind, with an R^2 of 0.82 and a higher MAE of 6.3, hence proving the low performance when it comes to handling such complex and dynamic pollution patterns in these areas. In fact, this capability of the proposed model to adapt to the changing sources of pollution in real time contributes to much to its effectiveness at such mixed-use recreational spaces.

Table 7: Overall Performance Summary Across All Regions

Method	Average R^2	Average MAE (AQI Units)	Average AQI prediction & healthcare enhancements efficiency (%)
Proposed	0.92	3.5	15.4
Method [4]	0.83	5.8	9.7
Method [9]	0.86	5.2	11.1
Method [15]	0.88	4.6	12.9

Overall, the general performance of the proposed model was better than that of Methods [4], [9], and [15] in all regions, averaging an R^2 of 0.92 and a mean absolute error of 3.5 AQI units. Indeed, the reduction in AQI by the proposed model, averaging 15.4% across regions, portrays its effectiveness in optimizing air quality in a wide range of environmental setups ranging from residential to industrial and traffic-heavy zones. Among these, Method [4] has the poorest performance w.r.t. both predictive accuracy and AQI prediction & healthcare enhancements efficiency, whereas Methods [9] and [15] perform moderately but fall behind the proposed model in handling the complex spatiotemporal dynamics of pollution. Further, we will discuss an example usage case of the proposed model that will help readers understand the whole process in more detail in different scenarios.

PRACTICAL USE CASE SCENARIO ANALYSIS

A multi-regional, systematic experiment was carried out to validate the proposed model and its sub-components in Delhi Geographies. Synthetic data, representative of the typical pollution, meteorological factors, and emissions from major sectors-generally residential and industrial sectors and heavy-traffic zones-were considered for the validations. The key features included pollutants like PM2.5, PM10, NO2, CO, SO2, and O3, besides meteorological indicators of temperature, humidity, wind speed, and wind direction. Feature selection, time series analysis, geospatial interpolation, optimization, and clustering analysis comprised the dataset pre-processing stages. Results in tabular form are elaborated in further sections about the performance at each component: XGBoost with Recursive Feature Elimination, ConvLSTM with Geospatial Interpolation, Multiple Agent Deep Q-Networks, and Dynamic Time Warping with DBSCAN. The results of this kind are analyzed with the motive of representation for predictive accuracy, spatial and temporal forecasting ability, optimization interventions, and clustering insights of a model in the process.

Table 8: XGBoost with Recursive Feature Elimination (RFE) – Feature Importance and AQI prediction & healthcare enhancements

Feature	Importance Score	Selected by RFE	Contribution to AQI (μg/m³)
PM2.5	0.89	Yes	54
PM10	0.82	Yes	42
NO2	0.76	Yes	30
CO	0.65	Yes	28
SO2	0.58	No	-
Temperature	0.50	Yes	10
Humidity	0.45	No	-
Wind Speed	0.35	No	-
Wind Direction	0.32	No	-
Atmospheric Pressure	0.40	No	-

In this regard, related to the selection of most important features for the prediction of AQI values, best results obtained by a combination of XGBoost with RFE on the dataset are depicted in Table 8. The features such as PM2.5, PM10, NO2, CO, and temperature were retained at each step of RFE because those features contributed most in that process towards AQI values. Eliminated features will include wind speed and direction since they have lower importance scores. Based on these features, the model has predicted AQI values, and their shares in the total AQI are respectively: PM2.5, PM10, and NO2 sets.

Table 9: ConvLSTM with Geospatial Interpolation (Kriging) – Temporal and Spatial AQI prediction & healthcare enhancements

Region	**Actual AQI (µg/m³)**	**Predicted AQI (µg/m³)**	**Spatial Error (µg/m³)**	**Temporal Accuracy (%)**
R. K. Puram	160	158	2	88.5
Okhla	220	215	5	90.1
Shadipur	140	135	5	87.6
Anand Vihar	180	175	5	85.0
Najafgarh	110	108	2	89.4
Mayapuri	200	195	5	92.0

Results of ConvLSTM combined with Kriging Table 9 presents the predicted AQI values of multiple regions showing temporal accuracy of the same. Spatial interpolation errors were small, in the range of 2-5 µg/m³, hence establishing the efficiency of the model in predicting AQI with good accuracy for those regions also where no monitoring stations are present. Temporal accuracy-the model's capability to capture pollutant trends overtime-exceeded 85% in all the regions, peaking at 92% in Mayapuri Geographies.

Table 10: Multiple Agent Deep Q-Networks (DQN) – Policy Optimization for AQI prediction & healthcare enhancements efficiency

Agent	**Action Selected**	**Reward Signal (Reduction in AQI)**	**Final AQI After Action (µg/m³)**	**Policy Improvement (%)**
Traffic Management	Vehicle Restriction	+18 µg/m³	160 -> 142	11.3
Industrial Emission	Emission Reduction	+25 µg/m³	220 -> 195	14.5
Municipal Regulation	Public Event Restriction	+15 µg/m³	180 -> 165	8.3
Pollution Control Board	Industrial Fine Increase	+20 µg/m³	200 -> 180	10.0
Traffic Management	Traffic Signal Adjustment	+12 µg/m³	140 -> 128	8.6

Table 10 represents various actions by the agents in the Multiple Agent DQN framework which try to reduce the AQI. Every action of the agents resulted in efficiencies within the AQI prediction & healthcare enhancements, while the reward signals became highest for industrial reduction of emission, thereby lowering AQI levels from 220 µg/m³ to 195 µg/m³ levels. The policy improvements, due to the dynamic optimization model, were as large as 8.3 to 14.5%, reflecting its ability to abate pollution in real time within and across sectors.

Table 11: Dynamic Time Warping (DTW) with DBSCAN – Clustering of Pollution Patterns

Cluster ID	Number of Regions	Average PM2.5 (µg/m³)	Average NO2 (µg/m³)	Average CO (µg/m³)	Outliers Detected
Cluster 1	5	95	45	0.80	2
Cluster 2	4	110	50	0.95	1
Cluster 3	3	150	70	1.20	3
Cluster 4	6	85	40	0.70	0

Table 11: DTW combined with DBSCAN on the clustering of regions into their pollution pattern series. There were four clusters, with Cluster 3 having the highest average of pollutant levels: PM2.5 is 150 µg/m³ and NO2 is 70 µg/m³ sets. Outliers from Clusters 1, 2, and 3 were found to be generally from industrial or high-traffic areas, proving the algorithm for applying it to detect abnormal patterns in pollution sets.

Table 12: Final Outputs – AQI prediction & healthcare enhancements and Reduction Summary Across All Regions

Region	Initial AQI (µg/m³)	Predicted AQI (µg/m³)	AQI prediction & healthcare enhancements efficiency (%)	Final AQI (µg/m³)
R. K. Puram	160	158	11.3	142
Okhla	220	215	14.5	195
Shadipur	140	135	8.6	128
Anand Vihar	180	175	8.3	165
Najafgarh	110	108	10.0	99
Mayapuri	200	195	10.0	180

Table 12: Summary of the overall AQI prediction & healthcare enhancements and reductions from model results across different regions after the full deployment of the model. The final AQI values were computed after implementing interventions identified by the Multiple Agent DQN framework. The best AQI prediction & healthcare enhancements efficiency, 14.5%, was realized in Okhla and reflected the effectiveness of industrial controls put in place. In traffic-congested regions, such as R. K. Puram and Shadipur could return AQI prediction & healthcare enhancements efficiencies of 11.3% and 8.6%, respectively, based on traffic congestion and optimization. From the results represented in Tables 8 to 12, the proficiency of the proposed model for all its components becomes evident. XGBoost with Recursive Feature Elimination identified major features contributing toward AQI, while ConvLSTM with Kriging estimated the spatial and temporal predictions with quite high accuracy even in areas absent of monitoring stations. MA DQN effectively optimized policies to reduce AQI values in real time, while DTW with DBSCAN presented clusters in the variation of pollution patterns that outlined the variation. Overall results summarized in Table 12 confirm efficiency for the model in providing a forecast on AQI values and, further interventions in real time can yield significant reductions in levels of pollution across several regions of Delhi Geographies. This multi-faceted approach underlines the practical utility of the model in urban air quality management, as it brings predictive accuracy together with actionable interventions in the process.

CONCLUSION AND FUTURE SCOPES

This section is dedicated to the presentation of an integrated comprehensive air quality prediction and optimization framework using XGBoost with RFE, ConvLSTM with Kriging, multiagent DQN, DTW with DBSCAN. The developed model outperforms other current methodologies in terms of its contribution to air quality improvement for various types of urban exposures, such as residential, industrial, mixed-use, and those with heavy traffic in Delhi. It has an average R^2 of 0.92 and a mean absolute error of 3.5 AQI units, hence outperforming the compared methods, with an average R^2 variation from 0.83 to 0.88 and MAE between 4.6 and 5.8 AQI units. Talking effectiveness with the model in real-time air quality management, the average enhancement in AQI levels stood in contrast to the reductions brought about by Methods [4], [9], and [15], which were 9.7%, 11.1%, and 12.9%, respectively. It performed even higher in specific regions. For instance, the developed framework ensured an AQI prediction & healthcare enhancements efficiency of 17.1% at traffic congestion hotspots like Pusa and ITO, confirming that it was able to adopt traffic mitigation strategies based on real-time pollutant levels. These results confirm, therefore, the efficiency of the integrated approach for capturing temporal-spatial dynamics of pollution and optimizing effective interventions to reduce its impact for the process. This framework thus ensures that machine learning and deep learning synergized with geospatial analysis methods will be perfectly suited for practical applications in an urban air quality management system.

FUTURE SCOPES

While the performance of the proposed framework has been extremely promising for various scenarios, there do remain some future areas of research and potential development. The dataset can be extended for longer temporal scales and more cities in India for a wider context to evaluate the generalizability of the model. This would also allow seasonal effects and long-term pollutant trends, thereby making the models robust for AQI forecasting over a very extended period of time, say months or years. The model can be extended to a more thorough simulation of secondary pollutants like VOCs and other aerosol parameters such as particulate sulfate and nitrate. This will enhance the accuracy of the model in predicting the formation of secondary pollutants, considered critical for regions experiencing rapid industrialization and vehicular growth. The integration of advanced reinforcement learning methods, such as multiple agent coordination algorithms that will permit agents to communicate and share policies, represents another potentially very promising direction. In this way, much more effective real-time optimization strategies can be enabled; this is crucial in scenarios where interventions in one region may have cascading effects on other adjacent regions-in this context, traffic management. Moreover, deploying it in currently operating real-to-life monitoring systems and developing interfaces toward real-time decision support would greatly enhance its usability for policy makers and urban planners. Integration with real-time feedback mechanisms using mobile sensors and low-cost air quality devices can further enhance model granularity and responsiveness, thereby rendering the models adaptive to dynamic urban environments. Introducing health impact metrics into the optimization framework may provide a more direct linkage between air quality improvements and public health outcomes, offering assurance that interventions reduce pollution but at the same time minimize adverse health effects among sensitive populations.

REFERENCES

[1] C. Liu, G. Pan, D. Song and H. Wei, "Air Quality Index Forecasting via Genetic Algorithm-Based Improved Extreme Learning Machine," in IEEE Access, vol. 11, pp. 67086-67097, 2023, doi: 10.1109/ACCESS.2023.3291146.
keywords: {Atmospheric modeling;Predictive models;Extreme learning machines;Prediction algorithms;Kernel;Air pollution;Genetic algorithms;Air quality;Extreme learning machines;Time series;air quality forecasting;machine learning;extreme learning machine;genetic algorithm},

[2] F. Farhadi, R. Palacin and P. Blythe, "Machine Learning for Transport Policy Interventions on Air Quality," in IEEE Access, vol. 11, pp. 43759-43777, 2023, doi: 10.1109/ACCESS.2023.3272662.
keywords: {Machine learning;Urban areas;Sensors;Atmospheric modeling;Observatories;Data models;Buildings;Transportation;Air quality;Air quality;clean air zone;data-driven framework;machine learning;policy validation;transportation system},

[3] S. Al-Eidi, F. Amsaad, O. Darwish, Y. Tashtoush, A. Alqahtani and N. Niveshitha, "Comparative Analysis Study for Air Quality Prediction in Smart Cities Using Regression Techniques," in IEEE Access, vol. 11, pp. 115140-115149, 2023, doi: 10.1109/ACCESS.2023.3323447.
keywords: {Air pollution;Atmospheric modeling;Predictive models;Prediction algorithms;Computational modeling;Regression tree analysis;Random forests;Machine learning;Air quality;Internet of Things;Smart cities;Air pollution;machine learning;IoT;smart city;air quality index},

[4] K. Chatterjee et al., "Future Air Quality Prediction Using Long Short-Term Memory Based on Hyper Heuristic Multiple Chain Model," in IEEE Access, vol. 12, pp. 123678-123693, 2024, doi: 10.1109/ACCESS.2024.3441109.
keywords: {Atmospheric modeling;Predictive models;Accuracy;Computational modeling;Air pollution;Data models;Prediction algorithms;Air quality;Heuristic algorithms;Air quality;air pollutant concentrations (APCs);deep learning (DL);heuristic;machine learning (ML);meteorological factors (MFs);multiple chain;regressors},

[5] M. Mehrabi, M. Scaioni and M. Previtali, "Forecasting Air Quality in Kiev During 2022 Military Conflict Using Sentinel 5P and Optimized Machine Learning," in IEEE Transactions on Geoscience and Remote Sensing, vol. 61, pp. 1-10, 2023, Art no. 4103310, doi: 10.1109/TGRS.2023.3292006.
keywords: {Atmospheric modeling;Air pollution;Predictive models;Prediction algorithms;Atmospheric measurements;Monitoring;Data models;Air quality monitoring;machine learning (ML);$PM_{2.5}$ concentration;sentinel 5P;Ukraine war},

[6] I. Mokhtari, W. Bechkit, H. Rivano and M. R. Yaici, "Uncertainty-Aware Deep Learning Architectures for Highly Dynamic Air Quality Prediction," in IEEE Access, vol. 9, pp. 14765-14778, 2021, doi: 10.1109/ACCESS.2021.3052429.
keywords: {Atmospheric modeling;Predictive models;Air pollution;Forecasting;Vehicle dynamics;Uncertainty;Deep learning;Conv-LSTM;spatio-temporel prediction;highly dynamic air quality;accidental pollutant release;uncertainty;FFT-07;WSN},

[7] K. Chatterjee et al., "Toward Cleaner Industries: Smart Cities' Impact on Predictive Air Quality Management," in IEEE Access, vol. 12, pp. 78895-78910, 2024, doi: 10.1109/ACCESS.2024.3406502.
keywords: {Atmospheric modeling;Predictive models;Data models;Spatiotemporal phenomena;Meteorology;Mathematical models;Air quality;Internet of Things;Smart cities;Smart grids;Atmospheric measurements;Weather forecasting;Air quality;air pollutant concentrations (APCs);Internet of Things (IoT);meteorological factors (MFs);smart city (SC);spatiotemporal;weather smart grid (WSG)},

[8] Y. Liu, J. Nie, X. Li, S. H. Ahmed, W. Y. B. Lim and C. Miao, "Federated Learning in the Sky: Aerial-Ground Air Quality Sensing Framework With UAV Swarms," in IEEE Internet of Things Journal, vol. 8, no. 12, pp. 9827-9837, 15 June15, 2021, doi: 10.1109/JIOT.2020.3021006.
keywords: {Monitoring;Sensors;Atmospheric modeling;Air quality;Data models;Unmanned aerial vehicles;Three-dimensional displays;Aerial-ground sensing framework;air quality index (AQI);computer vision;federated learning (FL);unmanned aerial vehicle (UAV)},

[9] Y. Cao, D. Zhang, S. Ding, W. Zhong and C. Yan, "A Hybrid Air Quality Prediction Model Based on Empirical Mode Decomposition," in Tsinghua Science and Technology, vol. 29, no. 1, pp. 99-111, February 2024, doi: 10.26599/TST.2022.9010060.
keywords: {Analytical models;Atmospheric modeling;Computational modeling;Time series analysis;Predictive models;Air quality;Air pollution;air quality prediction;Empirical Mode Decomposition (EMD);Singular Value Decomposition (SVD);AutoRegressive Integrated Moving Average (ARIMA)},

[10] P. Dey, S. Dev and B. S. Phelan, "Predicting Multivariate Air Pollution: A Gaussian-Mixture Nested Factorial Variational Autoencoder Approach," in IEEE Geoscience and Remote Sensing Letters, vol. 21, pp. 1-5, 2024, Art no. 1002805, doi: 10.1109/LGRS.2024.3416343.
keywords: {Air pollution;Atmospheric modeling;Predictive models;Long short term memory;Feature extraction;Forecasting;Data models;Air pollutant;deep learning (DL);factorial variational autoencoder;latent space;machine learning (ML)},

[11] Z. Wang, Y. Yang and S. Yue, "Air Quality Classification and Measurement Based on Double Output Vision Transformer," in IEEE Internet of Things Journal, vol. 9, no. 21, pp. 20975-20984, 1 Nov.1, 2022, doi: 10.1109/JIOT.2022.3176126.
keywords: {Air quality;Transformers;Monitoring;Feature extraction;Deep learning;Tensors;Atmospheric modeling;CNN;mobile devices;multihead self-attention;transformer;vision transformer},

[12] X. Liu et al., "Estimating Black Carbon Levels With Proxy Variables and Low-Cost Sensors," in IEEE Internet of Things Journal, vol. 11, no. 10, pp. 17577-17588, 15 May15, 2024, doi: 10.1109/JIOT.2024.3361977.
keywords: {Sensors;Pollution measurement;Calibration;Atmospheric modeling;Air quality;Atmospheric measurements;Machine learning;Air quality;black carbon (BC);low-cost sensor;machine learning;proxy},

[13] S. Ali, T. Glass, B. Parr, J. Potgieter and F. Alam, "Low Cost Sensor With IoT LoRaWAN Connectivity and Machine Learning-Based Calibration for Air Pollution Monitoring," in IEEE Transactions on Instrumentation and Measurement, vol. 70, pp. 1-11, 2021, Art no. 5500511, doi: 10.1109/TIM.2020.3034109.
keywords: {Monitoring;Air pollution;Calibration;Temperature sensors;Temperature measurement;Atmospheric measurements;Air pollution monitoring;air quality monitor (AQM);Internet of Things (IoT);long-range wide area network (LoRaWAN);low-cost sensor;machine learning;remote sensing;sensor calibration},

[14] K. Yadav, V. Arora, M. Kumar, S. N. Tripathi, V. M. Motghare and K. A. Rajput, "Few-Shot Calibration of Low-Cost Air Pollution (PM$_{2.5}$) Sensors Using Meta Learning," in IEEE Sensors Letters, vol. 6, no. 5, pp. 1-4, May 2022, Art no. 7001704, doi: 10.1109/LSENS.2022.3168291.
keywords: {Sensors;Calibration;Computational modeling;Data models;Monitoring;Training;Position measurement;Sensor signal processing;air quality;few-shot learning;low-cost sensor calibration;model-agnostic-meta learning (MAML);machine learning},

[15] J. Song, K. Han and M. E. J. Stettler, "Deep-MAPS: Machine-Learning-Based Mobile Air Pollution Sensing," in IEEE Internet of Things Journal, vol. 8, no. 9, pp. 7649-7660, 1 May1, 2021, doi: 10.1109/JIOT.2020.3041047.
keywords: {Sensors;Atmospheric modeling;Air pollution;Biological system modeling;Atmospheric measurements;Pollution measurement;Computational modeling;Air quality (AQ);big data;machine learning;ubiquitous sensing},

[16] C. Park, G. S. Kim, S. Park, S. Jung and J. Kim, "Multiple Agent Reinforcement Learning for Cooperative Air Transportation Services in City-Wide Autonomous Urban Air Mobility," in IEEE Transactions on Intelligent Vehicles, vol. 8, no. 8, pp. 4016-4030, Aug. 2023, doi: 10.1109/TIV.2023.3283235.
keywords: {Air transportation;Transportation;Atmospheric modeling;Training;Urban areas;Deep learning;Reinforcement learning;Urban-air-mobility (UAM);air transportation service;multiple agent deep reinforcement learning (MADRL);centralized training and distributed execution (CTDE)},

[17] J. Borah et al., "AiCareBreath: IoT-Enabled Location-Invariant Novel Unified Model for Predicting Air Pollutants to Avoid Related Respiratory Disease," in IEEE Internet of Things Journal, vol. 11, no. 8, pp. 14625-14633, 15 April15, 2024, doi: 10.1109/JIOT.2023.3342872.
keywords: {Atmospheric modeling;Predictive models;Air pollution;Machine learning;Deep learning;Data models;Time series analysis;Air pollution;light GBM;pyCaret;random forest (RF)},

[18] C. -T. Wu et al., "A Precision Health Service for Chronic Diseases: Development and Cohort Study Using Wearable Device, Machine Learning, and Deep Learning," in IEEE Journal of Translational Engineering in Health and Medicine, vol. 10, pp. 1-14, 2022, Art no. 2700414, doi: 10.1109/JTEHM.2022.3207825.
keywords: {Diseases;Predictive models;Medical services;Real-time systems;Hospitals;Wearable computers;Environmental factors;Precision health;artificial intelligence;wearable device;chronic obstructive pulmonary disease;panic disorder},

[19] A. Mena-Martinez, M. Davila Delgado, J. Alvarado-Uribe and H. G. Ceballos, "A Real-Life Evaluation of Supervised and Semi-Supervised Machine Learning Approaches for Indirect Estimation of Indoor Occupancy," in IEEE Access, vol. 12, pp. 118673-118693, 2024, doi: 10.1109/ACCESS.2024.3449810.
keywords: {Support vector machines;Temperature sensors;Estimation;Mathematical models;Hidden Markov models;Temperature measurement;Semisupervised learning;Environmental monitoring;Sensors;Occupancy estimation;semi-supervised learning;environmental sensors;machine learning},

[20] S. Yu and S. Shen, "Compaction Prediction for Asphalt Mixtures Using Wireless Sensor and Machine Learning Algorithms," in IEEE Transactions on Intelligent Transportation Systems, vol. 24, no. 1, pp. 778-786, Jan. 2023, doi: 10.1109/TITS.2022.3218692.
keywords: {Compaction;Asphalt;Support vector machines;Machine learning algorithms;Sensors;Mathematical models;Correlation;Asphalt mixture;compaction prediction;machine learning;particle characteristics;sensing technology},

[21] F. Naz et al., "Comparative Analysis of Deep Learning and Statistical Models for Air Pollutants Prediction in Urban Areas," in IEEE Access, vol. 11, pp. 64016-64025, 2023, doi: 10.1109/ACCESS.2023.3289153.
keywords: {Predictive models;Atmospheric modeling;Air pollution;Data models;Forecasting;Logic gates;Monitoring;Air quality;machine learning;deep learning;predictive models;statistical methods},

[22] G. D'Elia et al., "Influence of Concept Drift on Metrological Performance of Low-Cost NO2 Sensors," in IEEE Transactions on Instrumentation and Measurement, vol. 71, pp. 1-11, 2022, Art no. 1004811, doi: 10.1109/TIM.2022.3188028.
keywords: {Calibration;Monitoring;Instruments;Air quality;Training;Sensors;Uncertainty;Air quality monitoring;concept drift;instrument calibration;instrument maintenance;machine learning (ML);multiple linear regression (MLR);relative expanded uncertainty (REU)},

[23] N. Jin, Y. Zeng, K. Yan and Z. Ji, "Multivariate Air Quality Forecasting With Nested Long Short Term Memory Neural Network," in IEEE Transactions on Industrial Informatics, vol. 17, no. 12, pp. 8514-8522, Dec. 2021, doi: 10.1109/TII.2021.3065425.
keywords: {Forecasting;Neural networks;Predictive models;Long short term memory;Deep learning;Time series analysis;Biological system modeling;Discrete stationary wavelet transform (DSWT);multichannel neural network;multitask neural network;nested long short term memory (NLSTM)},

[24] S. Acharyya, S. Nag and P. K. Guha, "Selective Detection of VOCs With WO3 Nanoplates-Based Single Chemiresistive Sensor Device Using Machine Learning Algorithms," in IEEE Sensors Journal, vol. 21, no. 5, pp. 5771-5778, 1 March1, 2021, doi: 10.1109/JSEN.2020.3041322.
keywords: {Sensors;Gas detectors;Prediction algorithms;Temperature sensors;Machine learning algorithms;Classification algorithms;Chemical sensors;Indoor air quality;volatile organic compound;WO_3 nanoplates;hydrothermal method;selectivity;machine learning algorithm},

[25] M. C. J. Rodrigues, O. Postolache and F. Cercas, "Unobtrusive Cardio-Respiratory Assessment for Different Indoor Environmental Conditions," in IEEE Sensors Journal, vol. 22, no. 23, pp. 23243-23257, 1 Dec.1, 2022, doi: 10.1109/JSEN.2022.3207522.
keywords: {Heart rate variability;Humidity;Sensors;Thermal analysis;Temperature sensors;Temperature measurement;Electrocardiography;Ballistocardiography (BCG);heart rate variability (HRV);human thermal comfort;indoor air quality (IAQ);Internet of Things (IoT);machine learning (ML);photoplethysmography (PPG);respiration activity, scenarios},