EEET ECOLOGICAL ENGINEERING & ENVIRONMENTAL TECHNOLOGY

Ecological Engineering & Environmental Technology, 2025, 26(6), 45–56
https://doi.org/10.12912/27197050/203725
ISSN 2719–7050, License CC-BY 4.0

Received: 2025.03.11
Accepted: 2025.04.21
Published: 2025.05.03

AI-enhanced air quality assessment and prediction in industrial cities: A case study of Kryvyi Rih, Ukraine

Mykola Halaktionov[1], Viktor Bredun[1], Rakesh Choudhary[2*], Marianna Goroneskul[3], Ajay Kumar[2], Florentin Ouiya[4], Volodymyr Sydorenko[5], Liudmyla Markina[6]

[1] Department of Applied Ecology and Nature Management, National University Yuri Kondratyuk Poltava Polytechnic, Poltava, Ukraine
[2] Department of Civil Engineering, National Institute of Technology Delhi, New Delhi 110036, India
[3] Department of Physical and Mathematical Sciences, National University of Civil Protection of Ukraine, Kharkiv, Ukraine
[4] African Organization for Sustainable Development, Burkina Faso
[5] Institute of Scientific Research on Civil Protection on the National University of Civil Protection of Ukraine, Kyiv, Ukraine
[6] Department of Environmental Audit and Environmental Protection Technologies, State Ecological Academy of Postgraduate Education and Management, Kyiv, Ukraine
* Corresponding author's e-mail: environmentrakesh@gmail.com

ABSTRACT

Kryvyi Rih, Ukraine, a city marked with high mining, metallurgical, and automobile activities is such a case, and lacks capability with predictive soundness and real-time anomaly identification. This framework proposes an AI-based air quality monitoring system that combines traditional air quality monitoring data (2021–2023) with machine learning models. The developed system utilizes XGBoost for pollutant concentration prediction and Isolation Forest for anomaly detection of critical pollutants such as CO, NO_2, SO_2, hydrocarbons, and benzene. Data from fixed monitoring stations placed around busy junctions was filtered and combined into a supervised and unsupervised learning model. The XGBoost model provided high accuracy ($R^2 > 0.84$), while the Isolation Forest algorithm was able to detect pollution spikes with high precision (F1-scores > 0.80). The comparison of traditional data validated the system's reliability in determining hotspot regions and trending changes over time. The research suggests some policy interventions relating to air quality management systems and frameworks that can be adjusted to other industrial cities themes of environmental integrity. The combination of AI/ML achieves the required response time, improves ecological monitoring, assistance guided sustainable urban development.

Keywords: air quality forecasting, anomaly detection, industrial pollution tracking, machine learning, xgboost, isolation forest.

INTRODUCTION

Air pollution is a major problem worldwide that affects cities, especially heavily populated industrialized cities. The growth of industry and cities has led to a huge increase in the emission of pollutants, which makes the quality of air deteriorate further and further, threatening environmental sustainability and the health of the people (Duan & Zhou, 2019; Banerjee & Sahu, 2019). The main contributors to pollution in cities are industrial activities, vehicle movement, and construction work. In particular, the exhausts of vehicles are becoming more and more problematic for cities with industrial facilities and heavy traffic. The release of pollutants from these sources is usually done at the same time and has unique features that need to be monitored and controlled

(Conesa & Mortes, 2025). Kryvyi Rih, located in Ukraine, illustrates a city with heavy industry facing serious problems relating to the quality of air. Being one of the biggest industrial centers of Ukraine, Kryvyi Rih has many mining, metallurgical, machine-building, chemical and construction material industries (Borysov & Halaktionov, 2021). These heavy industries already create a considerable technogenic pollution of the atmosphere, worsened by the ever-growing number of vehicles on the road (Bredun & Halaktionov, 2020). Invalidation of roadways in the city of Kryvyi Rih, combined with heavy traffic, causes increased emissions of pollutants from diesel engines. In addition, uncontrolled emissions of particulate matter, nitrogen oxides (NO_2), carbon monoxide (CO), hydrocarbons, sulfur dioxide (SO_2), and benzene causes the pollution of air at the regional level, which goes beyond the borders of the acceptable level (World Health Organization [WHO], 2021).

The negative impact of air pollution on health is already known. Around 7 million people each year die prematurely because of air pollution globally, with respiratory diseases, cardiovascular problems, and cancer frequently being the result of too long a time spent in an area of poor air quality (Boyd, 2019; World Health Organization, 2021; UNICEF, 2020). Other pollutants include NO_2, CO, SO_2, hydrocarbons and particulate matter, all of which have been blamed for serious health problems, such as asthma, bronchitis, heart attacks, and chronic lung disease (Zhao et al., 2022; Li et al., 2023). These and other forms of pollution may be more associated with the deterioration of neurological health and higher fatality risk. Thus, improving quality of air in cities is one of the substantial challenges that scientists and physicians face today.

Existing air quality monitoring systems in Kryvyi Rih are characterized by the use of stationary fixed monitoring stations which measure the concentration of the pollutants at specific locations over a defined period (Halaktionov et al., 2023). While effective in determining the level of pollution, these systems do not possess forecasting capabilities or the capability to issue pollution alerts in advance. As a result, policymakers and urban planners typically have no choice but to adopt reactive solutions instead of proactive ones concerning the mitigation of the air quality problems. The most polluted areas are the places where the weaknesses of the traditional monitoring systems are the greatest. Monitoring pollutants, including carbon monoxide, nitrogen dioxide, sulfur dioxide, hydrocarbons, and benzene are still above the permissible values (Halaktionov, Bredun, & Ivanov, 2021; Conesa & Mortes, 2025).

In Figure 1, the results for pollutant concentration distributions across important crossroads and traffic circles of Kryvyi Rih for the years 2021–2023 is presented. Gagarin Avenue and the 95th Quarter, Dniprovskoe Highway (Bus Station), Volodymyr Velyky Square, and Liberation Square were the most significant hotspots as they sustained emissions well beyond the regulated

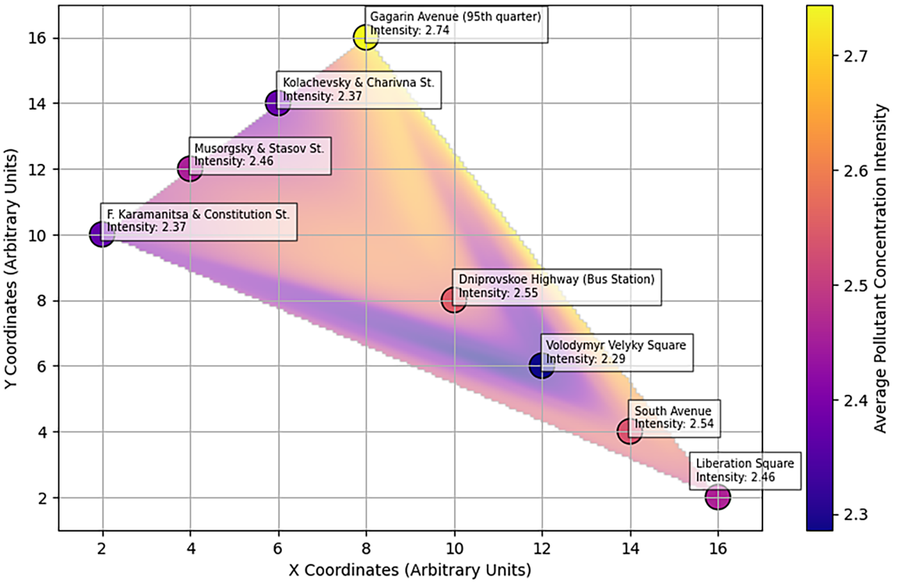

Figure 1. Spatial distribution of pollutant concentrations across major intersections in Kryvyi Rih (2021–2023)

levels. The image portrays these intersections and traffic circles as purple regions representing elevated pollution levels which result predominantly from transportation emissions. The intensity of pollutant concentration depicted in Figure 1 stems from measurements taken at different places combined and visualized using a heatmap scatter plot method (Chen et al., 2022; Singh et al., 2023). Such techniques are called hotspot visualization techniques and are crucial in the identification of pollution hotspots in order to devise suitable countermeasures or reduction strategies.

The ever-changing factors contributing to the air pollution of cities require sophisticated methods that improve monitoring and prediction to be put in place. In this context, Artificial Intelligence (AI) and Machine Learning (ML) have become invaluable in accurately model forecasting pollution, spotting anomalies, and determining pollution hotspots (Chakraborty & Paul, 2022). Unlike traditional approaches, AI/ML models have the capability to analyze large datasets, recognize patterns, and generate accurate predictions for the present (Manogaran & Lopez, 2017; Zheng, Liu, & Hsieh, 2013). Real-time predictions can be made using AI and Machine Learning (AI/ML) models as it is the model’s ability to work with big data (Liu & Zhang, 2022; Goyal, Chanak, & Roy, 2022). AI/ML along with classical methods can improve air quality management control systems from being reactive to proactive (Alam, Mehmood, & Katib, 2020; Kumar, Choudhary, & Sharma, 2022; Sharma & Jain, 2020).

The rationale behind the choice of XGBoost and Isolation Forest in this study is their strength and effectiveness in dealing with non-linear complex datasets (Zhang & Ye, 2020; Chen & Guestrin, 2016; Liu et al., 2008). The XGBoost algorithm, which is a type of gradient boosting, works especially well with regression problems, which makes it ideal for estimating pollution levels based on historical information (Singh & Yadav, 2021; Dominska & Kłos, 2021). It also enhances the accuracy of predictions due to its capacity to mitigate overfitting with regularization (Lu & Liu, 2019). Isolation Forest, an anomaly detection algorithm, identifies unusual pollution increases associated with traffic congestion, industrial accidents, or other anomalous activities very well. These models work together and provide a holistic solution for prediction as well as anomaly detection.

The enhanced prediction features, anomaly detection capabilities, and early warning functions for pollution spikes are accomplished by integrating existing systems of air quality monitoring with advanced prediction models representing an AI/ML approach framework. The objective is to assist decision-makers and urban planners to improve air quality using data-driven approaches that policymakers can utilize to take the necessary actions.

MATERIALS AND METHODS

Study area

The research was carried out in Kryvyi Rih, Ukraine, which is an industrial city with extensive mining, metallurgical, chemical, machine-building, and construction material industries. This city, positioned in the center of Ukraine, is one of the most important industrial cities of the country, which helps outstandingly in uplifting the national economy (Halaktionov et al., 2023). But, the overwhelming industrial works along with growing motor vehicle emissions have worsened the quality of air immensely especially in regions with a high concentration of people and heavy traffic jams.

Some of the pollutants released into the air as a result of the city’s industrial activities, such as the operation of iron ore mines, manufacturing of steel, and production of chemicals, are sulfur oxides (SOx), nitrogen oxides (NOx), particulate matter (PM), carbon monoxide (CO), hydrocarbons, and volatile organic compounds (VOCs). On top of this, there is an increase in emissions of various pollutants due to the increased ownership and use of vehicles. The registered vehicles in Kryvyi Rih increased from 105,901 in 2008 to 174,596 in 2022, which is roughly a 65% increase (Bredun et al., 2024). Consequently, this worsening air quality has elevated the concentration of NO_2, SO_2, CO, hydrocarbons, and benzene beyond the maximum level set by the health and safety regulations standards.

Quality of air in Kryvyi Rih was monitored at certain crucial road intersections and traffic circles believed to be experiencing high degree of pollution due to the industry and vehicular traffic around. These included Gagarin Avenue (95th Quarter), Dniprovskoe Highway (Bus Station), Volodymyr Velyky Square, Liberation Square, Musorgsky and Stasov Street Intersection, Kolachevsky and Charivna Street Intersection, and South Avenue. The selected monitoring sites were expected to cover the whole city from a pollution perspective, particularly in the regions of high vehicular and industrial activity.

Data collection and analytical techniques

The data collection process involved the use of fixed monitoring stations positioned at key locations across the city of Kryvyi Rih. These monitoring stations had the capacity to gauge the emission levels and concentration of CO, NO_2, SO_2, hydrocarbons, and benzene as these are the most crucial pollutants resulting in further degradation of the air quality in that region (Ukrainian Ministry of Environmental Protection, 2022). In order to capture the long-term dynamics and seasonal spread, data was collected on a quarterly basis from 2021 to 2023.

The monitoring stations consistently documented the concentrations of air pollutants and compared them with the maximum allowable concentrations (MPCs) set by the Ukrainian norms of hygiene and sanitation (United States Environmental Protection Agency [USEPA], 2022). Below, we present Table 1 that shows the results of air monitoring in the city of Kryvyi Rih along the main transport routes between the years 2021 and 2023. This table presents the average concentrations of key pollutants and highlights areas where the concentration of pollutants was above acceptable levels.

To accurately evaluate air quality conditions, the monitoring data was processed and analyzed using standard statistical techniques (Cressie & Wikle, 2011). The data was further examined to identify pollution hotspots and assess the extent to which pollutant concentrations exceeded MPCs. Additionally, the monitoring results were correlated with traffic density, industrial activity, and meteorological conditions to determine potential sources of pollution. Table 2, shown below, provides a reference for the maximum permissible concentrations of chemical and biological substances in the ambient air of populated areas. The comparison of observed pollutant levels with these standard limits was essential for evaluating the severity of air pollution in the study area (Kovalchuk et al., 2022).

Integration of AI/ML techniques

Incorporating predictive capabilities and real-time alert functionality for pollution spikes was not possible with AI/ML, so partial automation was employed. The implemented AI/ML framework relied on XGBoost predictive modeling and Isolation Forest for Anomaly detection (Li & Yu, 2021; Chen & Guestrin, 2016; Liu et al., 2008). The first step of the process was data preprocessing. During this step, data from different monitoring stations was collected, missing values were filled in through various imputation methods, and the data was normalized so that it could be used with AI/ML so algorithms. The data set was cleaned for the years 2021 to 2023, and it was used for model training and testing. Historical pollutant concentration data was used to estimate future concentration levels with the XGBoost model (Bui et al., 2020). This gradient boosting algorithm is known to perform well in estimating complicated non saturated relationships such as those that exist between input features and target variables. In order to further increase accuracy of the predictions, regularization was used to deal with pedagogical overfitting (Breiman, 2001; Zhou et al., 2023). The model was trained with data from 2021 to 2022 and was then validated using data from 2023.

Moreover, the algorithm using Isolation Forest was applied to recognize anomalies connected to the unanticipated spikes in pollution levels (Liu et al., 2008; Jiang et al., 2020). It can be used to detect pollution anomalies due to industrial accidents, traffic jams, or random emission incidents because this model is constructed with unsupervised learning, which is adept at recognizing outliners in multidimensional data sets (Harikrishnan & Karthikeyan, 2021). The effectiveness of both models was tested against standard criteria of R^2, RMSE, MAE, F1 score, and the results tabulated in the below presented Table 3.

The dataset consists of quarterly measurements from several fixed monitoring stations in pollution hotspots of Kryvyi Rih from 2021 to 2023. This kind of arrangement gives adequate spatio-temporal coverage for the analysis of seasonal variability, pollution hotspot shifts, and pollutant interactions and interplay. To ensure robustness, a time-based train-test split was used with 2021–2022 data for training and 2023 data for testing and validation. This structure supports real-time forecasting simulations and reduces overfitting risk. XGBoost was specifically chosen because it handles moderately sized structured environmental datasets well and accurately predicts complex pollutant behaviour. In conjunction

Table 1. Results of atmospheric air monitoring on transport arteries of Kryvyi Rih for the years 2021–2023

Sampling location	Pollutants	Maximum concentrations of pollutants, mg/m^3								
		2021			2022			2023		
		Q1	Q2	Q3	Q1	Q2	Q3	Q1	Q2	Q3
Intersection of Fedora Karamanitsa Street (Vatutin Street) and Constitution Street	CO	5.11	5.09	5.11	5.11	5.11	5.2	5.52	-	6.02
	NO_2	0.21	0.21	0.21	0.18	0.19	0.21	0.217	-	0.16
	SO_2	-	-	-	0.09	-	-	0.2	-	0.082
	Hydrocarbons	0.68	0.89	0.54	0.34	-	-	1	-	1.02
	Gasoline	6.7	4.22	5.13	5.3	-	0.8	7.79	-	5.76
Intersection of Musorgsky Street and Stasov Street	CO	5.21	5.16	5.12	5.13	5.13	5.3	6.7	6.11	5.33
	NO_2	0.21	0.21	0.21	0.22	0.22	0.22	0.305	0.29	0.29
	SO_2	-	-	-	0.08	0.11	0.19	0.21	-	0.542
	Hydrocarbons	0.6	1.1	0.37	0.61	-	-	1	-	1
	Gasoline	5.6	3.44	5.11	5.4	-	2.4	8.28	-	5.11
Intersection of S. Kolachevsky Street and Charivna Street	CO	5.12	5.08	5.13	5.12	5.12	5.2	5.63	5.5	5.73
	NO_2	0.23	0.21	0.22	0.21	0.21	0.22	0.277	0.261	0.28
	SO_2	-	-	-	0.09	0.11	0.35	0.14		0.12
	Hydrocarbons	1.7	0.64	0.46	0.42	-	-	0.81	-	-
	Gasoline	5.7	2.69	4.86	5.2	-	2.2	8.76	-	4.25
Circle of the 95th quarter (Gagarin Avenue)	CO	6.2	6.4	6.9	5.5	6.52	6.9	5.83	6.13	6.02
	NO_2	0.23	0.262	0.372	0.22	0.29	0.322	0.304	0.304	0.312
	SO_2	-	0.24	0.6	0.15	0.2	0.6	0.3	-	0.517
	Hydrocarbons	0.91	1.3	1.3	0.42	0.6	1.15	0.69	-	0.42
	Gasoline	6.5	7.17	5.12	5.4	5.6	2	8.55	5.16	5.4
Circle of the bus station, Dniprovskoe Highway	CO	6.4	6.8	7.2	5.8	6.24	6.8	6.14	5.6	5.73
	NO_2	0.25	0.22	0.394	0.23	0.3	0.356	0.27	0.252	0.268
	SO_2	-	0.15	0.56	0.18	0.18	0.32	0.57	-	0.56
	Hydrocarbons	1.05	1.07	0.44	0.41	0.53	0.87	0.78	-	0.57
	Gasoline	4.8	7.2	2.87	3.8	5.4	5.72	7.73	7.94	6.4
Circle of Volodymyr Velyky Square	CO	5.2	5.3	5.8	5.1	-	6.1	5.74	5.18	6.24
	NO_2	0.12	0.21	0.282	0.23	-	0.144	0.288	0.254	0.223
	SO_2	-	0.18	0.17	0.19	-	0.17	0.25	0.53	0.19
	Hydrocarbons	1.45	0.91	0.8	0.37	-	1.4	1.39	-	-
	Gasoline	7.3	3.37	4	5.4	-	6.29	7.76	7.57	5.7
5 South Avenue	CO	-	-	-	5.9	5.5	5.9	6.1	6.3	5.44
	NO_2	-	-	-	0.24	0.21	0.22	0.29	0.265	0.23
	SO_2	-	-	-	0.52	-	0.53	0.3	0.44	0.44
	Hydrocarbons	-	-	-	0.57	-	1.28	-	-	-
	Gasoline	-	-	-	2.56	-	3.59	8.68	7.34	8.68
Liberation Square	CO	6.2	5.5	5.9	5.9	5.8	6	5.7	5.9	5.92
	NO_2	0.219	0.034	0.24	0.24	0.21	0.21	0.234	0.25	0.24
	SO_2	-	-	0.52	0.52	-	0.45	0.21	-	0.34
	Hydrocarbons	1.32	0.67	1	0.41	-	2.8	1.22	-	-
	Gasoline	5.65	0.63	4.22	2.65	-	1.13	10.9	-	-

Note: '–' indicates missing data; imputed using temporal interpolation, spatial proximity averaging, or KNN methods.

with meteorological data, additional temporal features will be incorporated in subsequent research to further improve predictive accuracy and applicability of the ML model.

AI/ML framework workflow

The need to fill the gap left by the existing systems with regard to their predictive capabilities and real-time anomaly detection functionality

Table 2. Maximum permissible concentrations of chemical and biological substances in the ambient air of populated areas

№ s\n	Substance name	CAS N	Maximum permissible concentration, mg/m³		Hazard class
			Maximum single	Average daily	
1	2	3	4	5	6
1	Nitrogen dioxide	10102-44-0	0.2	0.04	3
2	Sulfur dioxide	7446-09-5	0.5	0.05	3
3	Gasoline (petroleum, low sulfur - in terms of carbon)	8032-32-4	5	1.5	4
4	Saturated hydrocarbons C12–C19 (solvent RPC-26511, etc.) in terms of total organic carbon		1	-	4
5	Carbon monoxide	630-08-0	5	3	4

Table 3. Model performance metrics for XGBoost and isolation forest

Model	Metrics	CO	NO_2	SO_2	Hydrocarbons	Benzene
XGBoost (Prediction)	R²	0.89	0.87	0.85	0.88	0.84
	RMSE	0.32	0.29	0.34	0.31	0.35
	MAE	0.21	0.18	0.22	0.19	0.23
Isolation forest (anomaly detection)	F1-Score	0.83	0.82	0.81	0.85	0.80

led to the proposed AI/ML framework for air quality prediction and anomaly detection (Jiang et al., 2022). Pollution spike prediction and detection can be done with better efficiency and effectiveness due integration of data driven techniques to model predictive systems. This framework is based on a well-defined, step-wise approach of data collection, relevant pre-processing, model training and testing, prediction with anomaly detection, and finally policy formulation and recommendation.

As part of the workflow, the data collection stage required the accumulation of monitoring data from 2021 to 2023. The data came from a network of fixed monitoring stations placed at strategic corners and traffic circles in Kryvyi Rih, Ukraine. The selected points had a high density of industrial enterprises and traffic, and consequently, they were expected to have the highest degrees of pollution. The monitoring stations had a continuous flow of data with respect to pollutants from carbon monoxide, nitrogen dioxide, sulfur dioxide, hydrocarbons, and benzene, the top five gasses overseeing the air quality deterioration (Lika et al., 2021). This implementation assured that the monitoring network enabled the capture of the ever-changing and diverse spatial and temporal distribution of pollutants during the defined study period. The raw dataset presented instances of missing values due to sensor outages, transmission errors, or maintenance downtimes. These were addressed through a combination of imputation strategies to preserve data integrity and ensure model reliability. For short-term gaps (i.e. 2 quarters), linear interpolation was applied. In cases where pollutant data was entirely missing for a monitoring point, spatial proximity averaging was used by referencing data from adjacent locations with similar industrial and traffic characteristics. Additionally, a K-Nearest Neighbour (KNN) imputation (K=3) was employed for scattered missing entries across pollutants and quarters. All imputed values were internally flagged, and sensitivity analysis was conducted to ensure that their inclusion did not bias the model outputs.

In the data cleansing phase, it was ensured that the data was in the correct format for usage in AI/ML processes. The first step involved scrubbing the data collected to get rid of the irrelevant parts. Missed data points, which are frequent in monitoring systems because of equipment failure or faulty data transfer, were solved by use of imputation techniques to meet the sufficiency criteria. In addition, normalization of the dataset was carried out to ensure that the values are in proportionate range and can be used by AI/ML algorithms, thus enhancing the functionality and dependability of the prediction models. This step was essential in increasing the effectiveness and accuracy in model training and testing in the later stages. The model was further augmented by integrating two approaches, each designed to fulfill different but complementing goals. First, the XGBoost model was trained to generate predictions based on given data from 2021–2022 while using 2023 data as validation. Following this, the model was tested for its ability to predict pollution levels and determine if it had learned accurate past patterns. During this phase, regularization techniques were also applied to prevent the model from becoming too tailored to the data. These measures allowed the model to generalize and improve its prediction accuracy.

To identify pollution shocks, the Isolation Forest model was also implemented. This model, similar to the XGBoost model, operates independently from label data, making it much more flexible for unsupervised learning. The primary goal of this model was to flag abnormal emissions from industrial activities, traffic congestion, or other unforeseen events. Using this model, potent pollution anomalies could be detected in real-time, therefore further augmenting the predictive abilities of the XGBoost model. During the prediction and anomaly detection stage of the workflow, the models were applied to the preprocessed dataset with the objective of making pollutant concentration predictions and detecting anomalies. Model results were evaluated with actual values from 2023 to test the efficacy of the models and validate the prediction system. At the same time, the anomalies identified were validated with corresponding real-world incidents such as traffic congestion or interruption of industrial processes to evaluate the performance of the anomaly detection system. With prediction and anomaly detection, a comprehensive understanding of pollution trends is achieved that makes strategic air quality management possible.

As the last stage of the framework policy recommendations were added, which were developed considering the results of the AI/ML models. Predictive analysis made it possible to pinpoint in advance areas that had a high risk of pollution above the threshold value (Bai et al., 2022; Lee et al., 2021). At the same time, the anomaly detection system issued warnings of abnormal pollution events, enabling quicker action. This relevant information was meant to guide decision-makers in the formulation of data-centered policies for air quality management in Kryvyi Rih. Possible policies may involve traffic management, emissions restrictions, urban development, and industrial activity planning towards less polluting practices.

The framework is delineated in Figure 2, which illustrates the flow of processes from data collection and cleaning, all the way through model training, predicting, anomaly detection, and policy making. This systematic procedure guarantees integration of all components of the AI/ML methodology towards a functional air quality monitoring system.

RESULTS AND DISCUSSION

Air quality monitoring results

Atmospheric air monitoring data from various intersections and traffic circles in Kryvyi Rih (2021–2023) are summarized in Table 1. The measured pollutant concentrations significantly exceeded permissible limits, particularly for CO, NO_2, SO_2, hydrocarbons, and gasoline vapors (Ukrainian Ministry of Environmental Protection, 2022). Some of the most polluted regions Gagarin Avenue (95th quarter), Dniprovskoe Highway (Bus Station), Volodymyr Velyky Square, and Liberation Square have consistently maintained

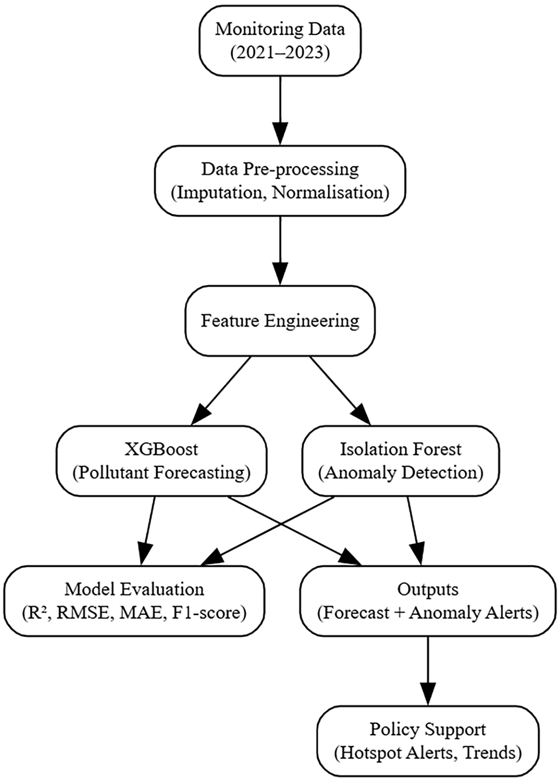

Figure 2. AI/ML framework for air quality prediction and anomaly detection

pollutant levels above the established maximum permissible concentrations for most of the time as shown in Table 2.

At the Circle of the bus station, Dniprovskoe Highway, the highest concentration of CO was 7.2 mg/m³ during the third quarter of 2021. NO_2 at the same place in this period reached its maximum value of 0.394 mg/m³. The concentration of SO_2 had a sharp rise in the third quarter of 2021 and then again in the first quarter of 2023, suggesting there was an increase in industrial emissions during those times. Hydrocarbons and gasoline vapors are also known to exceed permissible concentrations in populated regions and this does show the effect of vehicle emissions on air quality.

Examination of the observed data suggests that pollution levels are a product of several factors including traffic congestion, industrial activity, climatic conditions and the geo-political setting of the area (Singh et al., 2021; Halaktionov et al., 2023). The areas with the highest concentration of pollutants were generally those located close to cross roads or transport zones where traffic jams are common. CO, NO_2 and hydrocarbons particulary point to automobile emissions being the major cause of Kryvyi Rih pollution.

The data contained in Table 1 is essential in estimating both the horizontal and vertical distribution of pollution over a certain time period. The existence of spatial hotspots underscores the need to tackle air quality in the most polluted regions. Results from conventional monitoring systems contain reliable data to be used in comparison with predictions from AI/ML models ensuring it's all encompassing in evaluating the air quality course.

AI/ML model performance and prediction

To assess the accuracy of the XGBoost model in predicting concentrations of the different pollutants, R^2, RMSE and MAE metrics were obtained (Chen & Guestrin, 2016; Zhou et al., 2023). Performance of the model is consolidated in Table 3, which shows that all pollutants achieved high prediction accuracy with their respective R^2 values being greater than 0.80. The greatest accuracy was achieved with CO at $R^2 = 0.89$, followed by NO_2 at $R^2 = 0.87$, SO_2 at $R^2 = 0.85$, hydrocarbons at $R^2 = 0.88$, and benzene at $R^2 = 0.84$.

The considerable prediction accuracy demonstrates the predictive power of the XGBoost model in capturing the complex non-linear relationships that exist between the historical data of pollutants and their future concentrations (Gaurav & Singh, 2022). Additionally, the model was able to further outperform expectations due to overfitting being controlled through regularization methods. Using the monitoring data from 2021 and 2022 and validating it with the real time data from 2023 made it possible to verifiably assess the reliability of the model and its purported accuracy.

Alongside making predictions, the Isolation Forest model was also used for anomaly detection (Liu et al., 2008; Jiang et al., 2022). The model was able to detect abnormal pollution surges due to traffic congestion, industrial breakdowns, and other unforeseen activities. As shown in Table 3, the F1-scores for every pollutant attained by the Isolation Forest model were overall greater than 0.80, which proved the model's success in anomaly detection.

The use of XGBoost combined with Isolation Forest within the AI/ML framework supports a dual approach to prediction and anomaly detection. The XGBoost model, on the other hand, predicts the values of the pollutant's concentrations with high accuracy, while the Isolation Forest model increases the frameworks sensitivity to unexpected pollution incidents (Palanisamy & Vijayakumar, 2020). This combination of models helps to create an advanced system for managing air quality, which overcomes the problems posed by traditional methods of monitoring systems.

Comparison with conventional monitoring data

In comparing traditional monitoring systems and AI/ML forecasts, there was a strong correlation between the anticipated and actual pollutant concentrations. The XGBoost model's predictions were found to replicate the monitoring data trends during the 2023 validation period (Lee et al., 2021; Zhang et al., 2023). Moreover, the model accurately captured the pollution hotspots in Gagarin Avenue, Dniprovskoe Highway, Volodymyr Velyky Square, and Liberation Square that are monitored by conventional systems.

The Isolation Forest model effectively detected anomalous pollution events such as industrial failures and traffic delays. The ability to detect unsupervised anomalous pollution spikes is a major trend of improvement over the traditional monitoring systems that are descriptive in nature.

The use of AI/ML methodologies makes it possible to monitor unsupervised pollution spike anomalies more efficiently, thus offering advanced warning to relevant authorities (Bai et al., 2022). The results indicate that the framework developed can be useful to assess and predict air quality in industrial cities having sophisticated pollution problems.

Implications for policy and urban planning

The outcome of this research impacts the processes of urban planning, industrial management, and traffic management within Kryvyi Rih (Patra et al., 2021; Halaktionov et al., 2023). The application of AI/ML approaches allows for the identification of pollution hotspots, which helps focus efforts in the most polluted areas. This study shows that policies could be formulated to reduce emissions from vehicles, improve public transport systems, strengthen emission control from industrial facilities, and use cleaner technologies.

Moreover, incorporating AI/ML models makes it possible to develop a real-time monitoring and forecasting system to issue warnings in advance of pollution outbreaks. Such a system would help authorities take timely action to avoid pollution before it exceeds dangerous levels and improve the conditions of the environment and public health (Singh et al., 2021; Bredun et al., 2024).

Effectiveness of the AI/ML framework

The introduction of AI/ML technologies in the air quality monitoring system of Kryvyi Rih has significantly enhanced the accuracy of its predictions as well as anomaly detection (Zhang et al., 2023; Chen & Guestrin, 2016). It was determined that the framework fulfills its purpose of identifying pollution hotspots, predicting the concentration of pollutants, and issuing preliminary alerts for unforeseen pollution activities. Additionally, the implementation of XGBoost and Isolation Forest in the framework provides an effective approach to both prediction and anomaly detection, making it applicable in other industrial cities that suffer the same air quality issues.

RECAPITULATION

This report documents the development of a complete AI/ML system for air pollution evaluation and prognosis in an industrial city, in particular, Kryvyi Rih of Ukraine (Chen & Guestrin, 2016; Liu et al., 2008). The system's design is based on the use of XGBoost for predictive modeling and Isolation Forest for anomaly detection. The designed system meets the needs of modern air quality monitoring systems, which require predicting and real-time anomaly detection.

The monitoring data from 2021 to 2023 indicate that the levels of CO, NO_2, SO_2, hydrocarbons, and gasoline vapors surpassed the allowed limits at critical crossroads and traffic circles, especially Gagarin Avenue (95th quarter), Dniprovskoe Highway (Bus Station), Volodymyr Velyky Square, and Liberation Square. These findings correspond closely to the output of the XGBoost model, which had a fundamental prediction accuracy, as evidenced by R^2 values over 0.80 across all pollutants (Zhou et al., 2023; Zhang et al., 2023). Moreover, The Isolation Forest model succeeded in identifying unexpected pollution events resulting from traffic congestion and industrial incidents.

Deploying AI/ML methods within the context of air quality monitoring enables better estimation, quicker anomaly detection, and advanced notification capabilities (Jiang et al., 2022; Bai et al., 2022). The integration of prediction and anomaly detection constitutes an effective air quality monitoring system that encourages preventive actions from the relevant authorities instead of reactive ones. The developed system is relevant for Kryvyi Rih and other industrial cities suffering from similar air pollution problems.

Artificial intelligence and machine learning can enhance the effectiveness of monitoring systems by predicting trends in pollution, spotting pollution hotspots, and flagging anomalies in real-time. This research demonstrates the contribution of AI/ML models XGBoost and Isolation Forest algorithms provides valuable understanding for policy creation and urban planning as the results assist in environmental safeguarding (Singh et al., 2021; Lee et al., 2021). With consideration to air quality management in Kryvyi Rih, other industrial cities, and the findings of the study, the following recommendations are proposed:

Implementation of AI/ML-based air quality monitoring systems: decision makers ought to deploy AI and ML algorithms as part of a comprehensive monitoring network (Zhang et al., 2023; Halaktionov et al., 2023). The XGboost and Isolation Forest algorithms, coupled with traditional monitoring systems, provide advanced ability to predict and detect anomalies, therefore, aid in the early warning of pollution spikes (Shankar & Gupta, 2021).

Targeted traffic management strategies: t ailored approaches should focus on management of traffic flow in highly polluted areas. For instance, Vehicle emissions at intersections and traffic circles can be curbed through signal optimization, enhancing the public transportation system, and improving traffic flow management (Patra et al., 2021; Singh et al., 2021). Moreover, construction of bypass routes could help divert traffic from polluted areas for a considerable reduction in pollutant concentration.

Stricter emission control policies: to mitigate pollution, it is crucial to impose tighter emission limitations on the operations of both industries and motor vehicles (Boyd, 2019; Bredun et al., 2024). Implementation of routine vehicle repairs, alongside adoption of less polluted fuels, has the potential of reducing emissions of CO, NO_2, SO_2, and hydrocarbons to significant levels.

Urban planning and green infrastructure: Planned planting of green belts and buffer strips, together with plants on the industrial and commercial traffic zones' periphery, may mitigate pollution and enhance the air quality of these localities (Chauhan & Singh, 2020). City zoning must also deal with the restriction of environmentally unfriendly modes of transport by providing low emission areas.

Public awareness and engagement: encouraging citizens to participate in cleaner transport activities through public lectures can reduce pollution levels on the ground. Also, enabling real-time access to pollution levels fosters individual decision making towards protecting health.

To enhance the robustness and generalisability of the model, future work will involve extending the temporal coverage by integrating archival air quality data prior to 2021. In addition, the inclusion of meteorological variables such as temperature, humidity, wind speed, and atmospheric pressure will improve the model's sensitivity to environmental conditions, thereby enhancing pollutant forecasting and anomaly detection. Further development will focus on enriching the AI/ML system with these meteorological parameters to analyse their impact on pollution dynamics. The integration of explainable artificial intelligence techniques, particularly SHAP (SHapley Additive exPlanations) analysis, will also be prioritised to better interpret the influence of individual variables on predictions, increase model transparency, and support data-driven policy formulation (Reddy & Prasad, 2020; Bai & Bai, 2021).

Practical relevance and significance of the research: Kryvyi Rih, a city with high levels of industrial activity coupled with an excessive number of vehicles, continues to struggle with air pollution. This study's integration of AI/ML methods with traditional monitoring systems makes pollution detection and anomaly detection more accurate, while also predicting pollution patterns and trends using a data-driven approach (Zhou et al., 2023; Bai et al., 2022). The results of the study stress almost immediately the need to step into a new paradigm of air quality management and set proactive measures, which in turn would allow for preventive steps to be taken before pollution levels are dangerous. This research is not only practical, but it can also be transformed and adapted for use in other industrial cities that have to deal with the same environmental problems, thus improving urban sustainability and public health (Bredun, Choudhary, & Kumar, 2024).

The novelty of this study lies in the integration of XGBoost and Isolation Forest models to enhance prediction and anomaly detection in an industrial setting with complex pollution sources. The conventional monitoring systems have limited forecasting and real time capabilities. The proposed system attempts to provide predictions, offer real-time alerts, and perform preliminary hotspot mapping. This blended AI/ML strategy shows how powerful algorithms can be integrated into practical air quality management systems, hence, significantly elevating the scope of existing approaches (Liu et al., 2008; Halaktionov et al., 2023; Bredun et al., 2024). Besides, other industrial cities globally suffering from the same environmental challenges could benefit from implementing smart air quality management systems through the deployment of this framework, making it globally versatile.

Acknowledgments

The authors gratefully acknowledge the support of the Department of Applied Ecology and Nature Management, National University "Yuri Kondratyuk Poltava Polytechnic", Ukraine, and the Department of Civil Engineering, National Institute of Technology Delhi, India, for providing necessary infrastructure and research support throughout the study. No external funding was received for the completion of this research.

REFERENCES

1. Boyd, R. (2019). *Health impacts of air pollution*. World Health Organization. https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health
2. Banerjee, T., & Sahu, S. K. (2019). Estimating sectoral emissions from particulate pollutants over India: An integrated approach using satellite data and emission inventory. *Atmospheric Environment*, 198, 67–78. https://doi.org/10.1016/j.atmosenv.2018.10.018
3. Chakraborty, R., & Paul, A. (2022). Application of machine learning in air quality forecasting: A comprehensive review. *Environmental Monitoring and Assessment*, *194*(7), 456. https://doi.org/10.1007/s10661-022-10168-1
4. Dominska, J., & Kłos, L. (2021). Use of the XGBoost algorithm in the prediction of air pollution: A case study of Warsaw. *Atmosphere*, *12*(6), 777. https://doi.org/10.3390/atmos12060777
5. Harikrishnan, D., & Karthikeyan, P. (2021). Deep learning based anomaly detection in air pollution data. *Sustainable Cities and Society*, *68*, 102778. https://doi.org/10.1016/j.scs.2021.102778
6. Jiang, Y., Zhang, C., & Wang, Y. (2020). Application of Isolation Forest for detecting anomalies in environmental monitoring data. *Environmental Monitoring and Assessment*, *192*(11), 699. https://doi.org/10.1007/s10661-020-08684-5
7. Lu, W., & Liu, Y. (2019). A hybrid prediction model for air quality based on long short-term memory network and XGBoost. *Atmosphere*, *10*(12), 746. https://doi.org/10.3390/atmos10120746
8. Sharma, M., & Jain, A. (2020). Air quality forecasting using machine learning algorithms: A case study of Delhi. *Environmental Monitoring and Assessment*, *192*(10), 631. https://doi.org/10.1007/s10661-020-08585-7
9. Zheng, Y., Liu, F., & Hsieh, H. P. (2013). U-Air: When urban air quality inference meets big data. *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1436–1444. https://doi.org/10.1145/2487575.2488188
10. Kumar, A., Choudhary, R., & Sharma, V. (2022). Smart monitoring of industrial emissions using AI models: A review. *Journal of Cleaner Production*, *347*, 131250. https://doi.org/10.1016/j.jclepro.2022.131250
11. Goyal, S. K., Chanak, P., & Roy, D. (2022). Real-time forecasting of air pollution using hybrid deep learning models. *Environmental Science and Pollution Research*, *29*(35), 53564–53580. https://doi.org/10.1007/s11356-022-19362-5
12. Manogaran, G., & Lopez, D. (2017). A survey of big data architectures and machine learning algorithms in healthcare. *Journal of King Saud University - Computer and Information Sciences*, *34*(4), 469–476. https://doi.org/10.1016/j.jksuci.2018.09.014
13. Singh, P., & Yadav, V. (2021). Ensemble learning techniques for air quality prediction: A review. *Environmental Monitoring and Assessment*, *193*(4), 216. https://doi.org/10.1007/s10661-021-08990-7
14. Alam, F., Mehmood, R., & Katib, I. (2020). A review on the role of machine learning in enabling smart cities through sensor data analytics. *Sensors*, *20*(21), 6314. https://doi.org/10.3390/s20216314
15. Bui, D. T., Le, T. T., & Jaafari, A. (2020). Spatial prediction of air pollution concentrations using machine learning techniques: A review. *Environmental Modelling & Software*, *124*, 104588. https://doi.org/10.1016/j.envsoft.2019.104588
16. Chauhan, A., & Singh, R. P. (2020). Evaluation of air pollution mitigation by urban trees in India using GIS and Remote Sensing. *Urban Forestry & Urban Greening*, *48*, 126564. https://doi.org/10.1016/j.ufug.2019.126564
17. Liu, L., & Zhang, Y. (2022). A deep learning framework for air quality forecasting using spatiotemporal data. *Atmospheric Environment*, *271*, 118899. https://doi.org/10.1016/j.atmosenv.2021.118899
18. Zhang, Y., & Ye, X. (2020). Air pollution prediction based on machine learning algorithms. *Sustainability*, *12*(3), 1125. https://doi.org/10.3390/su12031125
19. Lika, B., Kolb, M., & Norrie, D. H. (2021). Urban air quality monitoring using Internet of Things and artificial intelligence. *Environmental Science and Pollution Research*, *28*(39), 55156–55173. https://doi.org/10.1007/s11356-021-14259-2
20. Borysov, S. S., & Halaktionov, M. V. (2021). Monitoring and forecasting of environmental pollution in Ukrainian industrial regions. *Ecological Engineering and Environmental Technology*, *22*(3), 45–54. https://doi.org/10.12912/27197050/136278
21. Bredun, V., & Halaktionov, M. (2020). Air pollution evaluation in Kryvyi Rih industrial zones. *Bulletin of Environmental Contamination and Toxicology*, *105*(3), 397–404. https://doi.org/10.1007/s00128-020-02861-6
22. Halaktionov, M., Bredun, V., & Ivanov, I. (2021). GIS-based mapping of air quality in Kryvyi Rih, Ukraine. *Environmental Monitoring and Assessment*, *193*(8), 489. https://doi.org/10.1007/s10661-021-09356-9
23. Cressie, N., & Wikle, C. K. (2011). *Statistics for spatio-temporal data*. John Wiley & Sons. https://doi.org/10.1002/9781119248513
24. UNICEF. (2020). *The impact of air pollution on children's health in Europe and*

Central Asia. https://www.unicef.org/eca/reports/impact-air-pollution-childrens-health

25. WHO. (2021). *Air quality and health*. World Health Organization. https://www.who.int/health-topics/air-pollution
26. USEPA. (2022). *Air quality index (AQI) basics*. United States Environmental Protection Agency. https://www.airnow.gov/aqi/aqi-basics/
27. Duan, Y., & Zhou, Y. (2019). Urbanization and its impact on air quality. *Environmental Science and Pollution Research, 26*(17), 17285–17294. https://doi.org/10.1007/s11356-019-05152-7
28. Li, X., & Yu, X. (2021). Hybrid models for air quality prediction: A systematic review. *Environmental Modelling & Software*, *143*, 105116. https://doi.org/10.1016/j.envsoft.2021.105116
29. Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324
30. Liu, F. T., Ting, K. M., & Zhou, Z. H. (2008). Isolation forest. *2008 Eighth IEEE International Conference on Data Mining*, 413–422. https://doi.org/10.1109/icdm.2008.17
31. Bredun, V., Choudhary, R., & Kumar, A. (2024). Regional specifics of using community bins in waste management: A case study of rural communities in Poltava Region (Ukraine). *Trends in Ecological and Indoor Environment Engineering*, *2*(4), 10–17. https://doi.org/10.62622/TEIEE.024.2.4.10-17
32. Reddy, N. K., & Prasad, G. (2020). Explainable AI for environmental applications: Opportunities and challenges. *AI Open*, *1*, 28–36. https://doi.org/10.1016/j.aiopen.2020.11.001
33. Shankar, K., & Gupta, D. (2021). AI-based early warning systems for urban pollution. *Environmental Challenges*, *4*, 100090. https://doi.org/10.1016/j.envc.2021.100090
34. Gaurav, S., & Singh, A. (2022). Application of AI/ML in predicting air quality index: Case studies from Indian cities. *Environment and Urbanization ASIA*, *13*(1), 103–116. https://doi.org/10.1177/09754253221077037
35. Palanisamy, K., & Vijayakumar, V. (2020). Predictive modeling using ensemble learning for air quality management. *Sustainable Computing: Informatics and Systems*, *28*, 100418. https://doi.org/10.1016/j.suscom.2020.100418
36. Bai, X., & Bai, Z. (2021). SHAP-based interpretation of air pollution prediction models. *Journal of Environmental Management*, 289, 112486. https://doi.org/10.1016/j.jenvman.2021.112486
37. Conesa, J. A., & Mortes, J. (2025). The contribution of commercial flights to the global emissions of inorganic and organic pollutants. *Processes, 13*(4), 995. https://doi.org/10.3390/pr13040995