Integrating ML and XAI with Urban Planning: Air Quality Predictions to Support Traffic Optimization

Marija Kopanja
BioSense Institute, University of Novi Sad
Novi Sad, Serbia
marija.kopanja@biosense.rs

Polina Maglevannaia
BioSense Institute, University of Novi Sad
Novi Sad, Serbia
polina.maglevannaia@biosense.rs

Marina Carević Tomić
Faculty of Technical Sciences
Novi Sad, Serbia
marinac@uns.ac.rs

Nikola Obrenović
BioSense Institute, University of Novi Sad
Novi Sad, Serbia
nikola.obrenovic@biosense.rs

Abstract

The increasing urbanization has intensified air pollution exposure, seriously threatening public health. This study examines how various machine learning and deep learning models can enhance the optimization of traffic changes within an urban planning framework, with a particular emphasis on air quality. Using local pollutant data from stations located in Novi Sad (Serbia), the models estimate emission changes from major transport infrastructure modifications. Two complementary approaches are used: (1) multi-output regression aimed at predicting air pollutant concentrations, and (2) multi-output classification focused on predicting the air quality index (AQI). For daily predictions, the Long Short-Term Memory (LSTM) model achieved the best $R^2 = 0.5$, while the Balanced Random Forest classifier reached 0.775 balanced accuracy for AQI classification. Finally, the most accurate models were applied to two prospective city network scenarios projected - (1) construction of two river bridges rerouting heavy traffic, and (2) creation of an extended pedestrian zone downtown - to quantify the expected reduction in pollutant concentrations and estimate possible changes in AQI categories. Depending on the pollutant, the results indicate a relative daily decrease between 0.7% and 2.9% on average for both scenarios. Although the improvement might seem small, this reduction is obtained for 13 street segments within a 100-meter radius around air stations, and a potential reduction for the whole city could be more substantial. Additionally, to turn forecasts into policy-ready evidence for traffic system decisions, we paired predictive models with explainable artificial intelligence (XAI) methods. Applying XAI methods reveals that meteorological factors, including relative humidity, total precipitation, temperature, and wind speed, primarily account for the predictions of developed models. Overall, the XAI methods highlight the combined influence of temporal, spatial, and meteorological factors on AQI.

ICAAI 2025, Manchester, UK

ACM ISBN 979-8-4007-2104-5

Keywords

Air Quality Prediction, Air Pollution, AQI, Machine Learning, LSTM, Explainable Artificial Intelligence, SHAP

ACM Reference Format:
Marija Kopanja, Polina Maglevannaia, Marina Carević Tomić, and Nikola Obrenović. 2025. Integrating ML and XAI with Urban Planning: Air Quality Predictions to Support Traffic Optimization. In *9th International Conference on Advances in Artificial Intelligence (ICAAI 2025), November 14–16, 2025, Manchester, UK*. ACM, New York, NY, USA, 7 pages.

1 Introduction and background

Rapid urbanization has significantly intensified air pollution exposure, posing a significant threat to public health, particularly affecting the respiratory and cardiovascular systems. According to the World Health Organization (WHO), air pollution causes approximately seven million premature deaths each year due to prolonged exposure to harmful pollutants such as particulate matter (PM), nitrogen oxides (NO_x), carbon monoxide (CO), and sulphur dioxide (SO2) [21]. Traffic-related emissions notably contribute to urban air pollution, as combustion engines directly emit pollutants including NOx, CO, and fine particulate matter (PM2.5, PM10), significantly impacting urban areas [6]. Accurate air quality predictions are thus essential for urban planners and local authorities aiming to mitigate pollution impacts and implement effective infrastructure modifications. Cities across Europe and beyond are integrating air quality considerations directly into urban planning initiatives such as traffic re-routing, low-emission zones, or pedestrianization schemes [3]. Studies have demonstrated that targeted traffic policy interventions can alter local air pollution dynamics and improve public health outcomes. For instance, the implementation of Low Emission Zones (LEZs) and congestion charging schemes in various European cities has resulted in significant reductions in air pollution-related health risks, especially cardiovascular diseases [1]. Similarly, in Dublin, targeted traffic pattern changes were associated with observable reductions in urban air pollutants, emphasizing the importance of spatially precise traffic management strategies for reducing pollution exposure [18].

Recent advancements in air quality prediction have increasingly utilized machine learning (ML) and deep learning methods to overcome limitations of traditional dispersion models, particularly regarding data scarcity and computational intensity. [16] utilized multivariate linear regression models to evaluate the effectiveness of

traffic restrictions in Padova, Italy, demonstrating significant correlations between traffic volume reductions and decreases in NO, NO_2, and NOx levels, although limited effects on PM10 were observed. Similarly, Wang et al. (2022) developed a prediction model combining Convolutional Neural Networks (CNN) with an improved Long Short-Term Memory (ILSTM) network, significantly enhancing predictive accuracy by effectively extracting relevant spatial and temporal features [20]. Additionally, [18] leveraged Gaussian Process Regression to capture the spatial variability of urban air pollutants in Dublin, illustrating the effectiveness of ML techniques in addressing data sparsity and spatial heterogeneity. Other studies [7, 13] have recognized the importance of traffic management systems in enhancing citizens' quality of life within smart city frameworks. Smart cities leverage information services to monitor public areas and infrastructure, which makes city services more aware, interactive, and efficient. Additionally, these studies proposed LSTM-based models for predicting Air Quality Index (AQI). The Temporal Sliding Long Short-Term Memory Extended Model (TS-LSTM) is proposed in [11] to predict AQI and concentrations for the following day or month, utilizing an optimal temporal lag to implement a sliding prediction window. Despite these advancements, critical gaps persist in current methodologies. Most models primarily focus on routine forecasts, providing limited insight into the potential impacts of proposed urban planning interventions, since many of them often operate as *black boxes*, offering minimal interpretability regarding the specific factors influencing air quality predictions.

Our study addresses these gaps by developing an integrated forecasting approach tailored specifically to the city of Novi Sad, Serbia. Novi Sad local authorities have proposed two significant urban infrastructure scenarios for 2024: (1) construction of two river bridges to reroute heavy traffic and (2) creation of an extended pedestrian zone downtown. Our study aims to contribute to understanding how these urban infrastructural changes would affect concentrations of pollutants and overall air quality. We leverage a comprehensive 15-year historical air quality dataset, integrating pollutant measurements (SO_2, NO_2, CO, NO, NO_x), ERA5 meteorological data, and graph-based road network metrics (Space Syntax). To estimate pollutant concentrations and air quality index (AQI) categories under these two scenarios, machine learning models are trained in two frameworks: multi-output regression and multiclass classification. Additionally, to convert forecasts into policy-ready evidence for traffic management decisions, we pair predictive models with explainable artificial intelligence (XAI) methods. Our findings contribute to optimizing the city's traffic network by providing robust support for urban planning decision-making processes. By using SHAP, we aim to determine the most important factors behind the decision-making process of complex black-box models in both frameworks. In recent years, many researchers have pointed out the lack of explanation methods for deep learning time series forecasting models [17]. SHAP method has been used in traffic [14] and pedestrian [8] accident analysis to obtain variable importance of different classification methods. In the study [22], it is shown that RF in synergy with SHAP can help in revealing the main factors influencing air pollutant measurements.

The remainder of the paper is structured as follows. A comprehensive description of the data is presented in the next section. The study design and methods used for multi-output regression and multi-class classification tasks are described in Section 3. The results of the analysis are presented and discussed in Section 4. Finally, the last section concludes the study with a summary of the contributions and limitations of this study, as well as future directions.

2 Data

2.1 Study area

The focus area of the case study is Novi Sad, Serbia (illustrated on Figure 1). The city of Novi Sad is located in the south of the Pannonian Plain at an elevation of 80-83 meters and covers an area of around $702,7 km^2$. The city is characterized by a temperate climate (by Köppen climate classification) and is exposed to a large amount of dust and strong winds, such as Košava, which can redistribute pollutants. The city area suffers from a high concentration of pollutants due to more intensive traffic, as well as other human activities, such as burning coal in industrial zones and households. For that reason, local authorities proposed two infrastructure plans. The first one, named the Novi Sad bypass, assumes building two new bridges as part of the city's outer ring, as visualized in Figure 1. The bypass is expected to relocate a significant amount of traffic flow from the west towards the east of the city and vice versa. The second plan is related to Petrovaradin, the oldest town within the Novi Sad area. To preserve the historical core of Petrovaradin, the main street (colored yellow on Figure 1) is planned to be closed to general traffic, while only public transport and delivery vehicles will still be allowed to enter. The street is a vital link between Petrovaradin and Novi Sad, and its closure could severely worsen traffic congestion. Therefore, evaluating this scenario is quite crucial for prospective stakeholders. To support effective urban planning, estimation of the potential reduction of pollutant emissions resulting from changes in transport infrastructure can help in the decision-making process in urban planning.

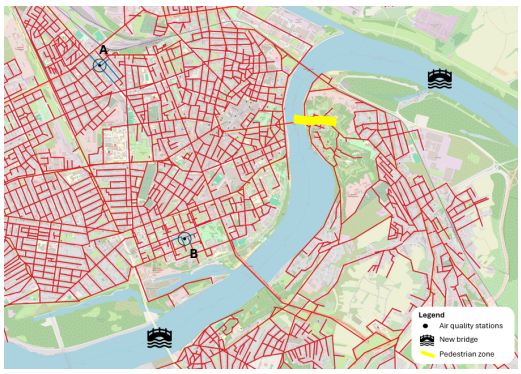

Figure 1: Study area and locations of air quality stations, A (north) and B (south). Red lines represent all street segments available in the data, and blue lines represent street segments within a 100-meter radius around the air quality station.

2.2 Air quality data

The Serbian Environmental Protection Agency (SEPA) operates two automatic air quality monitoring stations inside the study area (illustrated in Figure 1). Station A in the northern part of the city is located nearer the industrial zone and crossroads with heavy traffic. In contrast, station B in the southern part of the city is located in a park, not directly next to the main streets. Each observation contains the timestamp, the station identifier, station coordinates (latitude and longitude), and measures hourly mass concentrations of sulphur dioxide (SO_2), nitrogen dioxide (NO_2), nitrogen monoxide (NO), total nitrogen oxides (NO_x) and carbon monoxide (CO). The SO_2, NO_2, NO, and NO_x are expressed in $\mu g\,m^{-3}$, whereas CO is provided in $mg\,m^{-3}$. The raw files also contain air pressure P (mb), temperature t (°C) and relative humidity Rh (%), but these three fields are missing in roughly 60% of records and were therefore excluded from modeling. Although SEPA also reports hourly concentrations for ozone (O_3), fine particulate matter ($PM_{2.5}$), and inhalable particulate matter (PM_{10}), these variables suffer from substantial data gaps: only 43% of records for O_3 and $PM_{2.5}$, and 44% for PM_{10}, contain valid values. Given this high amount of missing values, these three pollutants were excluded from further analysis and model input.

To estimate the quality of air using SEPA data, categorically and numerically, the AQI parameter was introduced. In general, the standard set of pollutants used in Air Quality Index (AQI) calculation includes SO_2, NO_2, CO, $PM_{2.5}PM_{10}$, and O_3. Due to missing data for particulate matter and ozone in our case, we limited the AQI computation to the three pollutants with high coverage: SO_2, NO_2, and CO. The AQI used in our classification models was computed following the conventional breakpoint method. First, for each pollutant, a set of six concentration intervals $[C_{\min}, C_{\max}]$ and corresponding index ranges $[I_{\min}, I_{\max}]$ are defined. If a measured concentration C falls within one of these intervals, its sub-index I is obtained by linear interpolation (Equation (1)):

$$I = \frac{I_{\max} - I_{\min}}{C_{\max} - C_{\min}}(C - C_{\min}) + I_{\min}. \tag{1}$$

Once the sub-indices for all pollutants are calculated, the overall AQI is taken as the maximum sub-index (Equation (2)):

$$\text{AQI} = \max\{I_1,\ I_2, \ldots I_n\}. \tag{2}$$

Records for which one or more sub-indices could not be computed (due to missing pollutant concentration data) were assigned a missing AQI and excluded from classification. To convert the continuous AQI values into categorical labels, we initially used the standard six-category scale [19], which is redefined to four-class scheme shown in Table 1 due to severe class imbalance.

Table 1: AQI category definitions used in this study

Category	AQI	Description
AQI_1	$0 \le AQI \le 50$	Good
AQI_2	$51 \le AQI \le 100$	Moderate
AQI_3	$101 \le AQI \le 150$	Unhealthy for Sensitive Groups
AQI_4	$AQI > 150$	Unhealthy or Worse

2.3 Street network data

Space syntax techniques [12] and QGIS software [15] are used to analyze spatial layouts and human activity patterns in the study area. Using Space Syntax, the road geometry is decomposed into individual segments, meaning a single road in the original network may correspond to multiple segments in the processed network. The obtained network of Novi Sad for 2024 is shown in Figure 1. The network is analyzed using adapted graph-based metrics, such as betweenness centrality (choice), closeness centrality (integration) and others. The described features are calculated for each street segment; however, based on the locations of two air quality stations, only 13 street segments within a 100-meter radius (illustrated as blue circles around air quality stations on Figure 1) around the stations are considered relevant for measurements obtained on these stations. That is, the dataset includes 8 street segments located near station A and 5 segments near station B. The 100-meter radius is selected because the distance between air quality monitoring stations and roads can significantly influence the relationship between traffic and measured pollutant concentrations.

2.4 Meteorological data

The meteorological data, such as precipitation, wind and temperature, play significant roles in pollution formation. Therefore, we included meteorological parameters in our datasets. ERA5 is produced by the Copernicus Climate Change Service (C3S) at the European Centre for Medium-Range Weather Forecasts (ECMWF). ERA5 provides hourly estimates of a large number of atmospheric, land and oceanic climate variables. Variables used in our case study are 10m u-component of wind and 10m v-component of wind, representing horizontal speed of air moving towards the east and north, respectively, at a height of ten metres above the surface of the Earth. Furthermore, total precipitation, 2m dewpoint temperature (d2m), and 2m temperature (t2m) were also considered in our study. Relative humidity Rh (%) was recalculated from the t2m and d2m parameters, and replaces the incomplete SEPA Rh measurements. Daily values are computed by averaging all variables except total precipitation, which is summed from hourly data. These daily values are then aggregated into monthly data using mean values. Future work may exclude rainy days, as precipitation removes atmospheric particles, potentially introducing confounding effects in predictive modeling.

2.5 Datasets

From a spatial perspective, each datasets contains 13 street segments described with traffic and meteorological features, where the letter can be observed on a hourly, daily or monthly temporal scale. Historical data of in-situ measurements of multiple airborne pollutants, including SO2, NO2, CO, NO, and NOX, are accessible spanning 15 years, up to 2024. The only exception is the year 2022, for which pollutant data is not available. Descriptive statistic for daily datasets shows varying numbers of pollutant measurements within the same year range due to data scarcity. Some pollutants are missing for several months or entire years at a certain station. Average concentrations and standard deviations also vary across subsets and pollutants. For training, validation, and test set creation, a 10-4-1 schema is used to ensure sufficient training data

and a sizable validation set. Training data spans 2009–2018, while validation covers 2019, 2020, 2021, and 2023. For the test set, only one year, 2024, is used, to enable comparison with two additional scenarios related to urban changes in the city: (1) construction of two river bridges, rerouting heavy traffic and (2) creation of an extended pedestrian zone downtown.

Our modeling pipeline starts with integrating and cleaning heterogeneous data sources into unified tabular datasets. Hourly readings from two SEPA air quality stations were merged with street-segment attributes from Space Syntax and meteorological data from ERA5. Negative pollutant values were removed as physically invalid, and extreme outliers beyond realistic thresholds were excluded. Two dataset versions were created: one retaining samples with missing pollutant values, and another excluding them to assess their impact. Missing numerical inputs were imputed with means, and categorical features were one-hot encoded. Pollutant measurements were scaled consistently to address differences in scale. Temporal features such as year, month, weekday, and weekend indicators were added to capture seasonal and cyclical patterns. To capture diurnal patterns more explicitly, we also created three intraday subdivisions by splitting each 24-hour day into 3, 6, and 8 equal-length parts. This allowed the model to learn differences in air quality trends throughout the day.

The AQI class labels in the training, validation, and test datasets are not represented equally, a situation known as class imbalance. This imbalance can disrupt the process of generating machine learning classifiers, causing them to be biased towards the majority class. As a result, these classifiers may struggle to identify and accurately predict unhealthy levels in the minority class, which are typically of primary interest. The class imbalance in hourly and daily AQI distribution across training, validation and test sets could be observed in Figure 2.

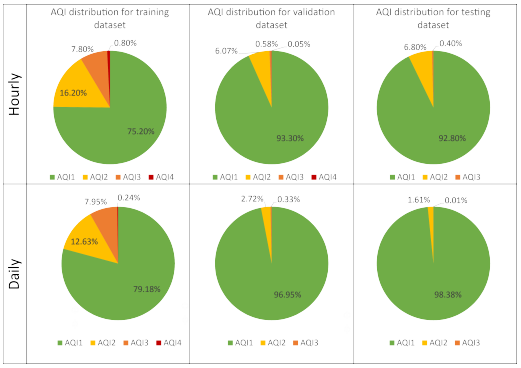

Figure 2: Hourly and daily distributions of AQI classes for training, validation, and testing datasets

3 Methodology

This study uses two complementary approaches for air quality prediction: (1) multi-output regression to estimate continuous pollutant levels, and (2) multi-class classification to predict AQI categories. A recurrent neural network was used for the spatio-temporal regression task, while various classifiers were applied to the classification problem.

3.1 Multi-output regression framework

In the multi-output regression framework, we are aiming to predict measurements of pollutants using a single model instead of having multiple models for each pollutant. Given the spatio-temporal nature of the data, long short-term memory (LSTM) networks [4] were used for their ability to capture long-term dependencies. Initial experiments with linear, polynomial, decision tree, and random forest regressors on daily and monthly datasets (Section 2.5) yielded poor performance ($R^2 < 0.2$), even with hyperparameter tuning, thereby highlighting the limitations of traditional models in handling high-dimensional and noisy sequential data. Accordingly, a deep learning network model based on LSTM layers was constructed, and a spatio-temporal modelling approach was used due to the temporal and spatial components present in the datasets. Input data sequences for the LSTM are constructed by preserving the spatial distinction between street segments, thereby enabling the model to capture both temporal dependencies and street segment-specific (spatial) characteristics. To account for time dependency and incorporate seasonal factors, the models were trained using a sliding window ranging from 7 to 30 days (with a 7-day increment). Additionally, shorter windows from 1 to 5 are also evaluated. The LSTM model is trained from all street segments jointly, where each sample is specified by the time window from a street segment, to learn a model with generalization ability across segments.

Numerous experiments have been conducted to identify the optimal hyperparameters for constructing the best LSTM structure. This has resulted in a network architecture featuring three main components. It begins with an input layer, followed by three hidden LSTM layers: the first and second each contain 128 units, while the third hidden LSTM layer consists of 64 units. The network is optimized using Adam algorithm, with mean absolute error (MAE) as the objective function due to its robustness on outliers (masked MAE loss function was used for dataset with missing values). The network is trained in 100 epochs with early stopping to prevent overfitting. The training is stopped if there is no improvement in MAE on the validation set after 10 epochs. The 3-layer stacked LSTM model is created, where each LSTM building block contains normalization layer and regularization layer with 0.2 dropout rate.

3.2 Multi-class classification framework

The multi-class classification framework addressed development of models on a granular classification scale involving four AQI categories. Initial experimentation with a standard Random Forest classifier provided baseline performance; however, the class imbalance issue, particularly in rare but critical high-pollution classes, persisted. To overcome this, we employed two specialized approaches: a multiclass LightGBM model [5] and a Balanced Random Forest classifier [2]. LightGBM has inherent capability to handle mild imbalance through class-weight adjustments. Balanced Random Forest method resamples training data at each bootstrap iteration to maintain class balance, thereby reducing bias toward majority categories without resorting to external oversampling.

All classification models were evaluated across multiple temporal resolutions, including hourly, daily, and intraday segments (3, 6, and 8 equal parts per day), to explore whether finer temporal segmentation could enhance predictive accuracy, particularly for

rapid shifts in air quality. The temporal split used mirrored that of the regression framework, with training data from before 2019, validation from 2019–2021 and 2023, and testing for the year 2024. Model hyperparameters were tuned using grid search guided by validation performance. For LightGBM, we optimized tree depth and learning rate, while Balanced Random Forest was optimized for number of estimators and minimum leaf size to improve minority-class sensitivity. Model performance was assessed using balanced accuracy (BA), which equally weighs recall across all classes.

Finally, after a comprehensive evaluation on our validation and test splits, the Balanced Random Forest classifier emerged as the most reliable model. We therefore selected this model to generate AQI forecasts for the year 2024 for two urban development scenarios: (1) scenario with two additional river bridges rerouting heavy traffic and (2) scenario with pedestrianization of zone downtown - to quantitatively estimate potential changes in AQI categories and their implications for urban planning and public health.

3.3 SHAP method

To enhance the interpretability of the most accurate models we employed SHapley Additive exPlanations (SHAP) [10] to identify which input features most strongly influenced AQI class predictions. For the Balanced Random Forest, as the best performer for the multi-class classification task and overall the best choice in framework, we used Tree SHAP, a model-specific XAI method for the exact calculation of SHAP values [9]. By aggregating the values, class-wise feature attributions are obtained on the 2024 test dataset. The SHAP values were computed for each AQI class, and both bar plots, showing global feature importance, and beeswarm plots, showing distribution of SHAP values per feature, were generated. This analysis was produced for each of the four AQI categories, allowing us to inspect the relative importance and effect direction of features such as temporal features, weather conditions, and street-segment descriptors.

4 Results and dicussion

4.1 Multi-output regression results

In the multi-output regression framework, the network is evaluated using mean squared error (MSE), root mean squared error (RMSE) and coefficient of determination (R^2). The performance of LSTM models was evaluated on two temporal resolutions—daily and monthly—using both the full and a filtered dataset.

For the daily scale, the 3-layered LSTM model (with a 7-day window) achieved moderate predictive performance, with slightly improved performance after removing missing values (filtered dataset). The visual inspection of results obtained from datasets with missing measurements (full dataset) for some pollutants indicated that the LSTM model produces predictions (mean values) that are overly smooth for those cases where true values are available. Overall, removing data with missing pollutant concentrations consistently led to improved performance across all metrics for both temporal scales, where the LSTM model benefited considerably on coarser monthly temporal resolution. Furthermore, the LSTM model performs better on daily compared to monthly data, probably due to much larger size of data, which is need for deep neural network models.

Table 2: Performance evaluation of LSTM models for daily and monthly scale.

	Full dataset			
	MAE	**MSE**	**RMSE**	**R2**
daily	8,606	289,758	17,02228	0,483
	Filtered dataset			
	MAE	**MSE**	**RMSE**	**R2**
	7,44	272,466	16,50654	0,508
	Full dataset			
	MAE	**MSE**	**RMSE**	**R2**
monthly	11,293	311,667	17,65409	0,181
	Filtered dataset			
	MAE	**MSE**	**RMSE**	**R2**
	8,198	171,932	13,11228	0,548

4.2 Multi-class classification results

To predict AQI categories across four classes, we compared the performance of Balanced Random Forest (BRF) and Light Gradient Boosting Machine (LGBM) classifiers.

Table 3: Balanced accuracy for multi-class AQI classification

	Balanced accuracy	
Temporal resolution	**BRF**	**LGBM**
Daily	0.7829	0.5000
8-hour windows	0.8120	0.5423
4-hour windows	0.5221	0.3650
3-hour windows	0.5137	0.3996
Hourly	0.5200	0.4914

As shown in Table 3, the BRF consistently outperformed LGBM across all resolutions, particularly on aggregated data. The best performance was observed for BRF on the 8-hour windows dataset with a BA of 0.812, closely followed by the daily model (0.7829). In contrast, LGBM performance remained considerably lower, with its best result of 0.542 also observed on the 8-hour windows dataset. This suggests that BRF is more robust to class imbalance and better captures AQI dynamics. Lower performance for both models was observed on finer-grained resolutions (e.g., hourly), likely due to high intra-day variability and smaller effective sample sizes per class, which challenge stable pattern learning in minority AQI categories.

To quantify the expected reduction in pollutant concentrations and estimate possible changes in AQI categories, the best models were applied to two scenarios: (1) construction of two river bridges rerouting heavy traffic and (2) creation of an extended pedestrian zone downtown. The results are shown in Table 4. Depending on the pollutant, the results indicate a relative daily decrease between 0.7% and 2.9% on average for both scenarios. With classification model, we estimated daily decrease from 3% up to 4.5%.

Table 4: Analysis of the impact of two urban infrastructure changes using the best models from both proposed frameworks.

	Multi-output regression			Multi-class classification	
Reduction (%)	***bridges***	***pedestrain zone***	***Reduction (%)***	***bridges***	***pedestrain zone***
SO2	0.34	0,1	AQI	3,15	4,46
NO2	2,34	0,73			
NOX	3,56	1,07			
CO	1,73	0,5			
NO	3,97	1,16			
Average	2,9	0,712			

4.3 Explanation framework

By using SHAP we tried to identify factors that affect the pollutant concentration and AQI level. Ultimately, our goal in developing an interpretable wrapper for both frameworks was to assist policymakers in creating urban plans that would positively impact air quality. Figure 3 shows a stacked bar plot where global attributions can be observed as well as per-class feature importance. Globally, the most important feature is relative humidity, followed by month and *SS node count*. Regarding meteorological parameters, the most influential variables, except relative humidity, are total precipitation and wind speed. Additionally, per-class analysis shows that the most important feature for the fourth AQI class is month, highlighting the significance of temporal components for the model to learn specific characteristics related to the most hazardous AQI category. From the aspect of infrastructural changes, the number of street segments along routes from a selected segment to others related to the traffic network is essential for predictions of AQI_1 and AQI_3 classes.

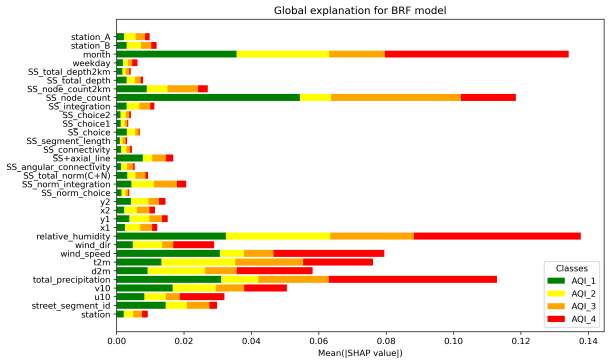

Figure 3: Explanation for Balanced Random Forest model for daily scale.

The typical bar feature importance plots do not tell whether some feature has a positive or negative effect on the model's decision. Therefore, to obtain this information, shap values are shown for each class separately on beeswarm plots.

The SHAP beeswarm plots of the BRF model's predictions show that the most influential features are month, relative humidity, and total precipitation, with smaller relative humidity generally associated with lower AQI values, suggesting that smaller humidity

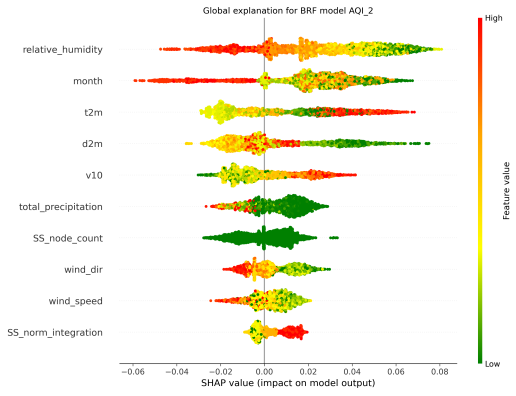

Figure 4: Explanation plot for AQI_2.

correlate with better air quality. Environmental variables such as precipitation, wind speed, and temperature also have notable impacts, where high precipitation and wind speed tend to have a positive impact on AQI, while high values of temperature tend to have an adverse effect on AQI. The month feature reveals seasonal patterns, with certain months contributing to higher pollution levels. Other features like v10, t2m, and *SS node count 2km* show moderate effects. Overall, the model highlights the combined influence of temporal, spatial, and meteorological factors on air quality. Similar conclusions can be drawn from Figure 5 obtained for LSTM model in multi-output regression framework.

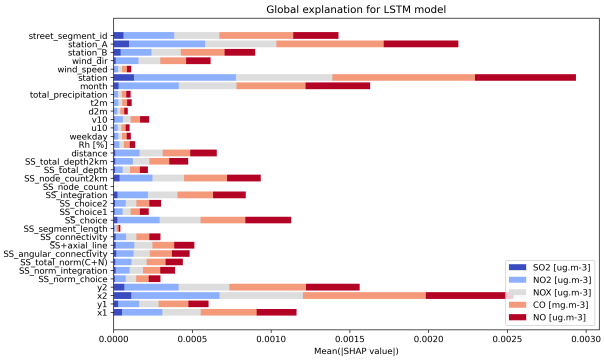

Figure 5: Explanation for LSTM model for daily scale (filtered dataset).

5 Concluding remarks

This study demonstrates the potential of machine learning and deep learning models to support urban planning decisions to improve air quality. By leveraging 15 years of historical data collected from various heterogeneous sources, including traffic network infrastructure, meteorological, and pollutant concentrations from the city of Novi Sad, Serbia, the models effectively captured patterns in pollutant concentrations and air quality index classifications. The LSTM model achieved the highest performance in daily-scale pollutant prediction, while the Balanced Random Forest classifier yielded the best results for AQI classification. These models were further

applied to simulate the impact of proposed urban infrastructure changes, projecting average daily reductions in pollutant levels between 0.7% and 2.9% in affected areas. Although modest, these results highlight the potential of AI-driven solutions to guide urban socio-economic activities and promote sustainable urban development. Importantly, integrating the SHAP method as an explanation tool provided insights into key contributing factors, such as precipitation, temperature, and wind speed, enhancing the transparency and policy relevance of the findings. However, the results indicate that data sources used may not be sufficient for creating effective predictive models, and other factors such as traffic flows should be considered. Overall, the approach offers a valuable framework for evidence-based decision-making in sustainable urban management.

Acknowledgments

This work has been carried out in the scope of the UDENE project, which has been funded by the European Commission in the scope of its Horizon Europe programme (contract number 101131190, https://udene.eu/).

References

[1] Rosemary C. Chamberlain, Daniela Fecht, Bethan Davies, and Anthony A. Laverty. 2023. Health effects of low emission and congestion charging zones: a systematic review. *The Lancet Public Health* 8 (2023), e559–e574. doi:10.1016/S2468-2667(23)00120-2

[2] Chao Chen and Leo Breiman. 2004. Using Random Forest to Learn Imbalanced Data. *University of California, Berkeley* (01 2004).

[3] European Environment Agency. 2024. Europe's Air Quality Status 2024. doi:10.2800/5970

[4] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long Short-Term Memory. *Neural Computation* 9 (11 1997), 1735–1780. doi:10.1162/neco.1997.9.8.1735

[5] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. 2017. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In *Advances in Neural Information Processing Systems*, I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Eds.), Vol. 30. Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2017/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf

[6] H. Khreis and et al. 2020. *Traffic-related air pollution: Emissions, human exposures, and health.* Elsevier.

[7] İbrahim Kök, Mehmet Ulvi Şimşek, and Suat Özdemir. 2017. A deep learning model for air quality prediction in smart cities. In *2017 IEEE International Conference on Big Data (Big Data).* 1983–1990. doi:10.1109/BigData.2017.8258144

[8] Sanghun Lee, Sangyeop Kim, Jaehoon Kim, Doyun Kim, Dohyun Lee, Gwangmuk Im, Yuk Hyeonseop, and Tae-Young Heo. 2023. Multiclass Classification by Various Machine Learning Algorithms and Interpretation of the Risk Factors of Pedestrian Accidents Using Explainable AI. *Mathematical Problems in Engineering* 2023 (05 2023). doi:10.1155/2023/1956865

[9] S.M. Lundberg, G. Erion, H. Chen, A. DeGrave, J.M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, and S. Lee. 2020. From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence* 2, 1 (2020), 2522–5839.

[10] Scott Lundberg and Su-In Lee. 2017. A Unified Approach to Interpreting Model Predictions. arXiv:1705.07874 [cs.AI] https://arxiv.org/abs/1705.07874

[11] Wenjing Mao, Weilin Wang, Limin Jiao, Suli Zhao, and Anbao Liu. 2021. Modeling air quality prediction using a deep learning approach: Method optimization and evaluation. *Sustainable Cities and Society* 65 (2021), 102567. doi:10.1016/j.scs.2020.102567

[12] Akkelies Nes. 2014. *Space Syntax in Theory and Practice.* 237–257. doi:10.1007/978-3-319-08299-8_15

[13] Esteban Pardo and Norberto Malpica. 2017. Air Quality Forecasting in Madrid Using Long Short-Term Memory Networks. 232–239. doi:10.1007/978-3-319-59773-7_24

[14] Amir Bahador Parsa, Ali Movahedi, Homa Taghipour, Sybil Derrible, and Abolfazl (Kouros) Mohammadian. 2020. Toward safer highways, application of XGBoost and SHAP for real-time accident detection and feature analysis. *Accident Analysis & Prevention* 136 (2020), 105405. doi:10.1016/j.aap.2019.105405

[15] QGIS Development Team. 2021. *QGIS Geographic Information System.* QGIS Association. https://www.qgis.org

[16] Riccardo Rossi, Riccardo Ceccato, and Massimiliano Gastaldi. 2020. Effect of Road Traffic on Air Pollution. Experimental Evidence from COVID-19 Lockdown. *Sustainability* 12, 21 (2020). doi:10.3390/su12218984

[17] Divyani Sen. 2025. Explainable Deep Learning for Time Series Analysis: Integrating SHAP and LIME in LSTM-Based Models. *Journal of Information Systems Engineering and Management* 10 (03 2025), 412–423. doi:10.52783/jisem.v10i16s.2627

[18] Pavlos Tafidis, Mehdi Gholamnia, Payam Sajadi, Sruthi Krishnan Vijayakrishnan, and Francesco Pilla. 2024. Evaluating the impact of urban traffic patterns on air pollution emissions in Dublin: a regression model using Google Project Air View data and traffic data. *European Transport Research Review* 16 (2024), 47. doi:10.1186/s12544-024-00671-z

[19] U.S. Environmental Protection Agency. 2024. *Technical Assistance Document for the Reporting of Daily Air Quality – the Air Quality Index (AQI).* EPA-454/B-24-002. Office of Air Quality Planning and Standards, Air Quality Assessment Division, Research Triangle Park, NC.

[20] Jingyang Wang, Xiaolei Li, Lukai Jin, Jiazheng Li, Qiuhong Sun, and Haiyao Wang. 2022. An air quality index prediction model based on CNN-ILSTM. *Scientific Reports* 12 (2022). doi:10.1038/s41598-022-12355-6

[21] World Health Organization. 2024. Ambient (outdoor) air pollution. https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health Accessed: 2024-05-14.

[22] Zhongcheng Zhang, Bo Xu, Weiman Xu, Feng Wang, Jie Gao, Yue Li, Mei Li, Yinchang Feng, and Guoliang Shi. 2022. Machine learning combined with the PMF model reveal the synergistic effects of sources and meteorological factors on PM2.5 pollution. *Environmental Research* 212 (2022), 113322. doi:10.1016/j.envres.2022.113322