28-oji Lietuvos jaunųjų mokslininkų konferencijos „Mokslas – Lietuvos ateitis" teminė konferencija
Proceedings of the 28th Conference for Junior Researchers "Science – Future of Lithuania"

DARNI APLINKA / SUSTAINABLE ENVIRONMENT

2025 m. balandžio 24 d., Vilnius, Lietuva
24 April 2025, Vilnius, Lithuania

ISSN 2029-7157 / eISSN 2029-7149
eISBN 978-609-476-390-8
https://doi.org/10.3846/da.2025.017

Geodezija ir kadastras / Geodesy and Cadastre *https://vilniustech.lt/346238*

ASSESSING AI ALGORITHMS FOR PREDICTIVE MODELLING OF SPATIOTEMPORAL PM_{10} AIR POLLUTION

Mina Adel Shokry FAHIM[1*], Jūratė SUŽIEDELYTĖ VISOCKIENĖ[1], Raimondas GRUBLIAUSKAS[2]

[1]*Department of Geodesy and Cadastre, Environmental Engineering Faculty, Vilnius Gediminas Technical University, Vilnius, Lithuania*
[2]*Department of Environmental Protection and Water Engineering, Vilnius Gediminas Technical University, Vilnius, Lithuania*

**E-mail: mina-adel-shokry.fahim@vilniustech.lt*

Received 24 February 2025; accepted 17 March, 2025

Abstract. Without a doubt, air pollution is one of the most serious issues confronting our world today, which presents significant health and environmental risks, exacerbating respiratory ailments and contributing to climate change. Air pollutants' spatial and temporal variability is the basis for effective air quality management, necessitating more accurate predictive models. The study aims to assess particulate matter of a diameter smaller than 10 µm (PM_{10}) forecasts using the European Union's Space Copernicus program mission of monitoring the atmosphere and tracking air pollutants, the Sentinel-5 Precursor satellite (5P) TROPOspheric Monitoring Instrument (TROPOMI), coupled with meteorological variables and observations from air quality monitoring stations. Root mean square error (RMSE) and mean absolute error (MAE) measure the model's accuracy. The study integrated machine learning algorithms and diverse datasets to enable precise spatial modelling of PM_{10} concentrations using a geographic information system (GIS). The results obtained peak accuracy during the heating season validation yielded an RMSE of 4.52 $\mu g/m^3$, MSE of 20.44 $(\mu g/m^3)^2$, and MAE of 3.30 $\mu g/m^3$, while testing resulted in an RMSE of 4.38 $\mu g/m^3$, MSE of 19.21 $(\mu g/m^3)^2$, and MAE of 3.19 $\mu g/m^3$.

Keywords: GIS, machine learning, meteorological data, particulate matter, predictive accuracy, remote sensing, Sentinel-5P TROPOMI.

1. Introduction

The process of transitioning to a greener world is unquestionably accompanied by the persistence of air pollution control strategies; particulate matter (PM_{10}) is one of the most critical pollutants, as it impacts human and respiratory systems, as well as the environment. Pollutant levels were traditionally extracted using a ground-based network of monitoring stations, which are typically limited in regional scope and tend to overlook air quality status in a global or regional context. However, satellite technology has advanced to the point where it is now possible to quantify air pollution over broad spatial and temporal ranges. While the use of artificial intelligence (AI) and its application is an undeniable aspect of technology use, this study collected data for PM_{10} observation along monitoring stations in Lithuania from January 2021 to December 2023. It assessed the data using machine learning and GIS techniques, coupled with spatial and temporal integration with remote sensing data for column densities of nitrogen dioxide (NO_2), carbon monoxide (CO), and ozone (O_3), as well as aerosols absorbing aerosol index (AOI) with meteorological variables. The study created assessment indicators such as RMSE, MAE, and MSE and then modelled the results on a two-dimensional map.

This research aims to provide comprehensive and precise air quality assessments across Lithuania. Its contributions are summarised as follows: the study employs the Tropospheric Monitoring Instrument (TROPOMI), a satellite instrument on board the Sentinel-5P. Secondly, it assesses the effectiveness of the proposed method for using ensemble learning algorithms for forecasting PM_{10} levels.

2. Literature review

Various studies have been conducted on pollution mapping, and numerous scientific papers have produced air pollution maps. The necessity of utilising advanced technology for prediction mapping has been established. The TROPOMI offers column densities for various pollutants, making its use essential to examine the air pollution levels in Pune, India, from 2020 to 2023 (Shah et al., 2024). It monitors pollutants like CO, Formaldehyde (HCHO), NO2, and others using satellite data from Sentinel-5P TROPOMI and ground station data. The findings show fluctuations in pollution levels over time, especially higher rises seen in the post-lockdown phase, suggesting the resumption of human activity. The application of machine learning is almost a necessary element of our civilisation. Another study was carried out in Siliguri city, also in India, using Random Forest among other machine learning methods (Das & Sahu, 2024). It uses satellite-derived Aerosol Optical Depth (AOD) data, meteorological and land cover criteria, to evaluate the effectiveness of several regression models. The random forest model was quite predictive, with a coefficient of determination R-squared error (R2) score of 0.83. Scientist Ketu uses the Central Pollution Control Board of India's data to show how machine learning techniques for air quality prediction, especially about AQI and NOx concentrations, are implemented (Ketu, 2022). Concerning alternative models, the research shows that the linear regression-based recursive feature elimination with random forest regression (RFERF) model demonstrates improved accuracy and predictive capability. In Croatia, Mamić et al. (2023) investigate the use of open-source remote sensing data to predict $PM_{2.5}$ and PM_{10} concentrationsClick or tap here to enter text.. Using data from Sentinel-5P TROPOMI, the study uses a random forest method to show that these models can estimate air quality with moderate to high accuracy. It focuses on variations in heating and non-heating seasons and regional variances in air pollution. The Mediterranean area achieved the best performance on PM10 during the heating season, with an R^2 value of 0.73. For Spain, forecasts PM_{10} levels in the key industrial and transportation node via integrating inverse distance weighting (IDW) interpolation with artificial neural networks (ANNs) (Rodríguez-García et al., 2024). Short-term PM_{10} concentrations are predicted using ANN models; comprehensive geographic forecasting maps are produced using IDW.

3. Materials and methods

This section will present the methods and materials used to obtain the results. This section is subdivided into four subsections, i.e., data acquisition description and pre-processing of data, proposed methodology, accuracy indicators, and GIS Modelling. In Section 1, the comprehensive discussion about the data acquisition dataset has been conferred and focuses on the cleaning and integration steps involved in data pre-processing. In Section 2, we explore the proposed modelling techniques. Section 3 details the accuracy indicators used to evaluate the model's performance. In Section 4, the study covers and applies GIS modelling and interpolation techniques.

3.1. Data acquisition and pre-processing

The dataset used in this research spans 2021 through 2023 for both heating and non-heating seasons. While the non-heating season is dedicated to June, July, and August, the heating season is set for December, January, and February. The dataset includes observations from the Sentinel-5P satellite's TROPOMI sensor. Google Earth Engine (GEE) was used to compile these Copernicus program data. There were around 2548 observations post-cleaning and integration for non-heating season data, and 2433 records for heating season data. Table 1 shows all input variables used to estimate surface PM_{10} across three years at 11 spatially scattered monitoring stations, under Lithuania's "National Air Monitoring Network". Stations with unique codes monitor environmental data in Vilnius, Panevėžys, Šiauliai, Klaipėda, Kaunas, and Kėdainiai municipalities.

Table 1. Input variables used to estimate surface PM_{10}

Dataset	Variable	Type	Data source
Monitoring Stations	Surface PM_{10} concentrations	Numeric	EEA*
TROPOMI	NO_2 column density, CO column density, O_3 column density, AOI absorbing aerosol index	Numeric	Sentinel-5p
Metrological conditions	Air Temperature, Feels Like Temperature, Wind Speed, Wind Gust, Wind Direction, Cloud Cover, Sea Level Pressure, Relative Humidity, Precipitation	Numeric	Lithuanian Hydro-meteoro-logical Service
Station Code		Numeric	
City Id		Numeric	
Date time serial		Numeric	

*European Environment Agency (EEA)

The study amalgamates many environmental and atmospheric data sources to provide a thorough perspective on air quality and meteorological conditions. The data encompasses surface PM_{10} concentrations from 11 monitoring stations, obtained from the EEA (European Environment Agency, n.d.) across diverse regions. The dataset employs the TROPOMI aboard the Sentinel-5P satellite to record atmospheric compositions (Ecosystem, 2025), including NO2, CO, and O3 total column densities (mol/m2) and the absorbing AOI. The Lithuanian Hydrometeorological Service's meteorological data, such as air temperature, wind speed, wind direction, cloud cover, and relative humidity, comprise identifiers for each city.

We prepare and improve the data for practical analysis at the stage of preparation and integration. Using Python's pandas, the timestamps in these datasets are standardised to UTC TimeStamp to ensure temporal consistency between datasets. The datasets are temporally integrated with the corrected datetime, linking air pollution data to the relevant meteorological conditions at exact times and locations. The entries were removed due to data quality indicators. "Validity" and "Verification" show uncertain data. Using an interquartile range (IQR) filtering approach for station observation values, cleaning operations included eliminating rows with missing data and outliers. First and third quartiles (Q1 and Q3) and the interquartile range (IQR) for these values were computed; thereafter, any values over 1.5 times the IQR below Q1 or above Q3 were excluded (Saradjian & Akhoondzadeh, 2011). The cleaned dataset is further categorised by time and location, aggregating pollutant concentrations and meteorological information into daily averages.

3.2. Methodology

Researchers have occasionally effectively used several machine learning techniques for long-term and short-term air quality prediction. Thus, we empirically tested machine learning ensemble approaches across AQI datasets and determined the performance of the bagged trees (BT) algorithm model. Figure 1 illustrates the flowchart of the study methodology.

A methodical strategy for examining PM_{10} pollution by integrating diverse data sets. Data is initially gathered from three primary sources: the Sentinel-5P satellite for atmospheric composition, local meteorological data for environmental conditions, and ground-based monitoring stations for localised PM_{10} concentrations. The datasets undergo pre-processing to guarantee correctness and consistency, including data cleaning and format standardisation. Thereafter, the data is temporally and spatially integrated, aligning all measurements to uniform geographic locations and TimeStamps for thorough analysis. The machine learning BT algorithm was utilised on the combined dataset to model and forecast PM_{10} levels. The models for both heating and non-heating seasons are validated and tested to verify their efficacy in forecasting pollution levels. Following that display by GIS modelling, which illustrates the distribution of PM_{10} pollution across.

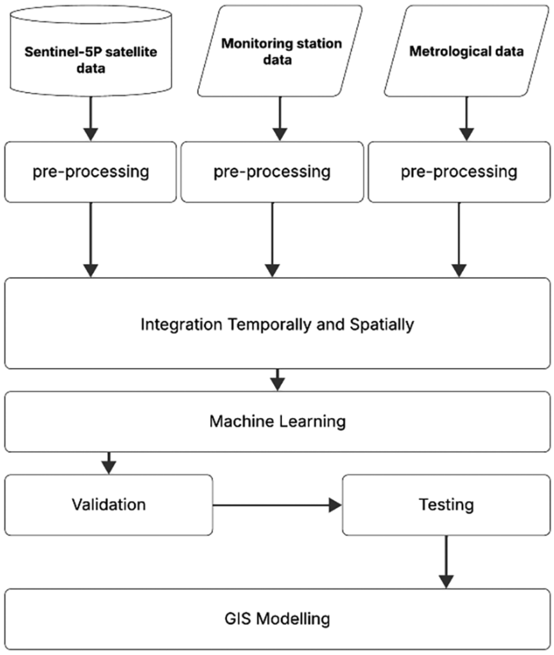

Figure 1. The flowchart of the study methodology

After pre-processing, subsequent steps might be employed to prepare data for testing and training. Studies in deep learning use several strategies to process data and develop models. This research intends to employ the BT algorithm, which is robust and adept at capturing intricate nonlinear interactions within datasets. The origins of the BT algorithm are detailed in Equation (1).

$$f(x) = \frac{1}{B}\sum_{b=1}^{B} T_b(x), \qquad (1)$$

where $f(x)$ represents the averaged prediction from B decision trees, $T_b(x)$ is the prediction from the b^{th} tree in the ensemble.

The BT algorithm uses an ensemble approach, wherein numerous decision trees are generated from various information segments, and their outputs are subsequently averaged. BT is quite good at addressing variance and overfitting problems, which are common obstacles in interpreting environmental data (Bbeiman, 1996; Breslow & Aha, 1997).

3.3. Accuracy indicators

The accuracy of machine learning models can be evaluated using various criteria. However, the mean absolute error (MAE) and the RMSE are commonly utilised as evaluation measures for measuring air pollution (Ayturan & Ayturan, n.d.; Hu et al., 2017; Larkin et al., 2017; Mamić et al., 2023; Rybarczyk & Zalakeviciute, 2018). After training and model development, 5-fold cross-validation was used to assess model performance. A key model evaluation metric is the MSE for each fold in Equation (2). The test MSE is calculated by averaging the MSE values from all folds. This method divides the dataset into five parts. Each iteration uses one subset for validation and four for training. This ensures that each data point is used once for training and validation, improving model dependability in Equation (3) (Bates et al., 2022; Stanford CS Theory, n.d.).

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \widehat{y_i}\right)^2, \tag{2}$$

where n is the number of observations in the fold, y_i is the actual value of the i^{th} observation, and $\hat{y}_i$ is the predicted value by the model for the i^{th} observation.

$$\text{Test MSE} = \frac{1}{k}\sum_{i=1}^{k}\mathrm{MSE}_i, \tag{3}$$

where k represents the number of folds and the MSE of the i^{th} fold.

MAE has been employed in "Environmental studies and air quality management", as it enables a straightforward and robust comparison of model predictions with the observations, as illustrated in Equation (4) (Chen et al., 2023; Lin et al., 2022).

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \widehat{y_i}\right|. \tag{4}$$

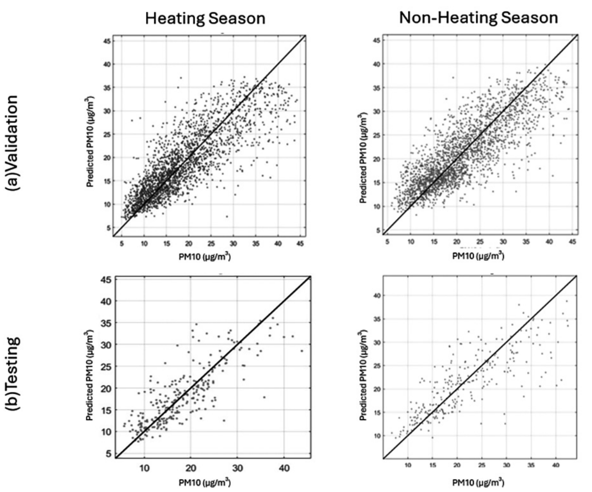

Figure 2. Validation and testing of model forecasts against observed PM_{10} levels

Ten percent has been allocated for testing purposes; the separate test set helps to evaluate the model's performance objectively in real-world events not covered by the training data. Residual plots can assess the efficacy of the models, validate predicted versus actual plots, and compare the BT algorithm for heating and non-heating seasons. Figure 2 shows how closely the model's forecasts align with the observed variables. Precision of the models is evaluated in the validation and testing stages. Every phase is executed as validation, shown in Figure 2 (a), and testing, shown in Figure 2 (b), displaying scatter plots of predicted against actual values (Tredennick et al., 2021).

3.4. GIS modelling

The interpolation maps on a national scale provide a visual representation of the seasonal variations of PM_{10} across the entire country and a visual distinction between the values of *in situ* stations and predicted data (Mamić et al., 2023). Spatially aggregating data based on the stations and calculating each mean value to highlight the discrepancies for each station between the *in situ* actual value and the predicted one, as illustrated in Figures 3 and 4.

IDW is implemented, the approach that is frequently implemented in GIS to approximate values at undetermined locations by utilising the values of known sample points. Figure 5 illustrates that prediction maps are comparable to those produced using *in situ* data. This method is based on the premise that the influence of a known point decreases as it approaches the undetermined location. The distance increased to the power parameter p is inversely proportional to the weight allocated to each known location. The weight of each sample point is determined as the inverse of its distance to the unknown point increased to a power parameter (p), which is derived using Equation 5, whereas the (p) is the power parameter regulates the importance of proximate points compared to remote ones. An increased p-value assigns greater significance to proximate points, yielding a more localised interpolation (Li & Heap, 2011; Lu & Wong, 2008; Shepard, n.d.).

$$w_i = \frac{1}{d_i^p}, \tag{5}$$

where w_i is the weight of the i^{th} point, d_i is the distance between the i^{th} point and the unknown point, and p is the power parameter.

The interpolation formula for the estimated value $\hat{Z}(x_0)$ at an unknown location x_0 is calculated using Equation 6.

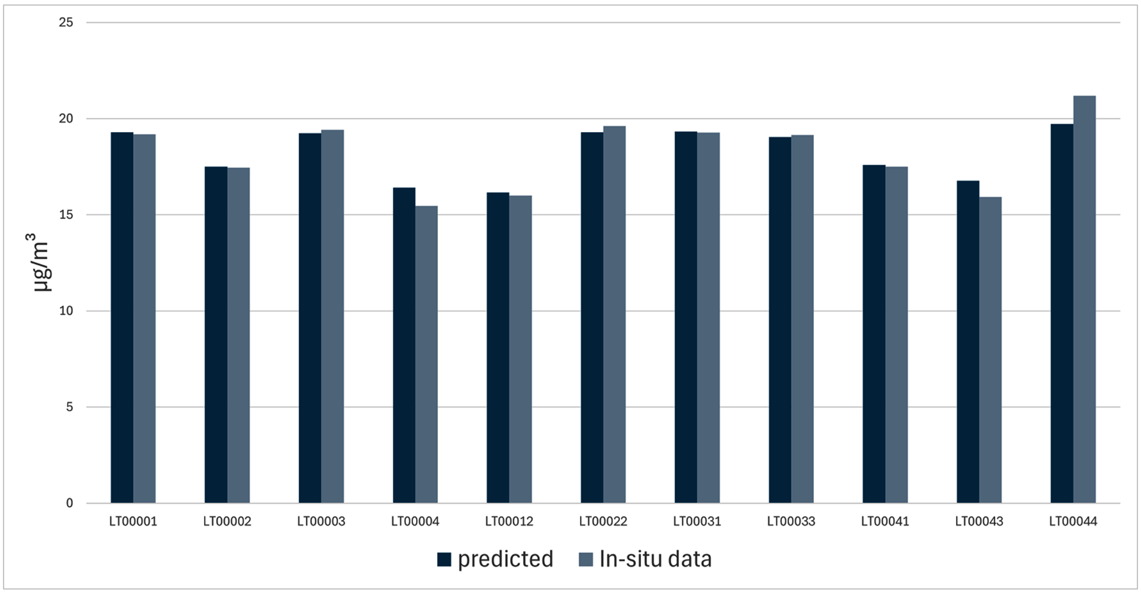

Figure 3. Heating season PM_{10} between *in situ* and predicted values

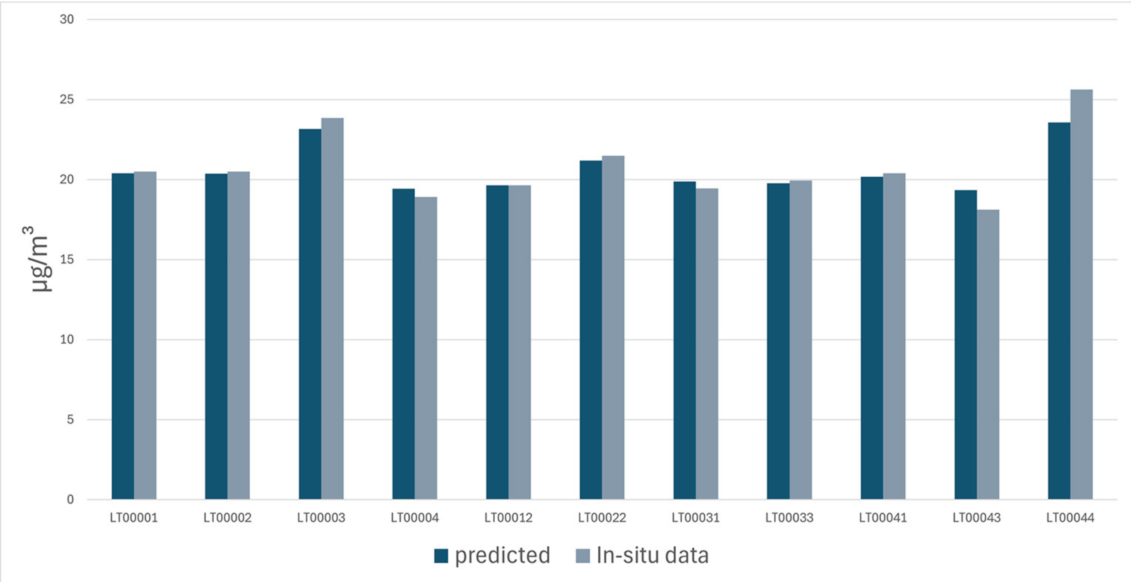

Figure 4. Non-heating season PM_{10} between *in situ* and predicted values

$$\hat{Z}(x_0) = \frac{\sum_{i=1}^{n} w_i Z(x_i)}{\sum_{i=1}^{n} w_i}, \quad (6)$$

where $\hat{Z}(x_i)$ is the value at the i^{th} sample point, and n is the number of sample points used in the interpolation.

4. Results and discussion

The study's results are defined by three key accuracy metrics of RMSE, MSE, and MAE, which were assessed throughout the validation and testing phases for non-heating and heating seasons. In the non-heating season, model validation reveals an RMSE of 4.52 μg/m^3, an MSE of 20.47 (μg/m^3)^2, and an MAE of 3.41μg/m^3, signifying consistent model performance with moderate forecast inaccuracies. Testing during the same season yielded modestly greater errors, with an RMSE of 4.97 μg/m^3, an MSE of 24.72 (μg/m^3)^2, and an MAE of 3.58 μg/m^3 as depicted in Table 2.

Table 2. Performance of models

		RMSE, μg/m^3	MSE, (μg/m^3)^2	MAE, μg/m^3
non-heating season	Validation	4.52	20.47	3.41
	Test	4.97	24.72	3.58
heating season	Validation	4.52	20.44	3.30
	Test	4.38	19.21	3.19

In contrast, the validation results for the heating season indicate an RMSE of 4.52 μg/m³, an MSE of 20.44 (μg/m³)², and an MAE of 3.30 μg/m³, closely resembling the validation results of the non-heating season. Testing findings throughout the heating season demonstrated enhanced accuracy, with an RMSE of 4.38 μg/m³, MSE of 19.21 (μg/m³)², and MAE of 3.19 μg/m³. In the non-heating season, the RMSE, MSE, and MAE rose by about 9.96 %, 20.76 %, and 4.99 %. For the Heating Season, these measures dropped by about −3.10 % for RMSE, −6.03 % for MSE, and −3.33 % for MAE. The results indicate that the models exhibit marginally greater accuracy in forecasting PM_{10} levels during the heating season relative to the non-heating season.

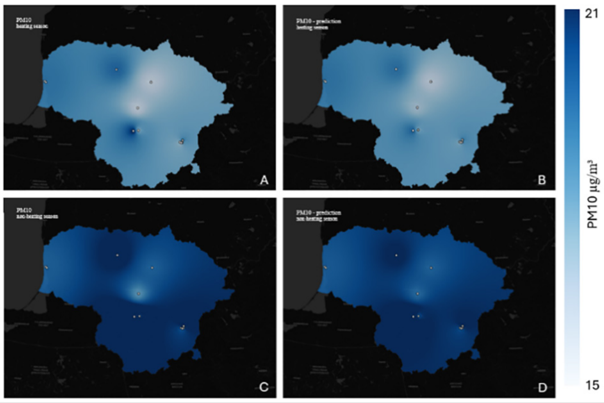

Figure 5. Interpolated PM_{10} concentration model for Lithuania: A – heating season from in situ data, B – heating season from predicted data, C – non-heating season from in situ data, and D –non-heating season from predicted data

The incorporation of TROPOMI data improved the analysis of trends in the relevant areas. GIS modelling interpolation reveals that PM_{10} concentrations during the heating season are considerably lower than those in the non-heating season, indicating superior air quality over the period. Figure 5 presents a comparative analysis of observed and predicted PM_{10} levels across Lithuania during heating and non-heating seasons. A and C display the actual PM_{10} concentrations measured at various monitoring stations during the heating and non-heating seasons. B and D depict the predicted PM_{10} concentrations for the same periods, generated by the machine learning model described in the study. Assessing a direct visual evaluation of the model's predictive accuracy by contrasting observed data against predictions across different seasonal impacts on air quality.

5. Conclusions

The research illustrates the prediction of spatiotemporal PM_{10} air pollution utilising AI algorithms, including the period from 2021 to 2023 for both heating (December, January, February) and non-heating (June, July, August) seasons, together with meteorological and satellite data. The integration of Sentinel-5P TROPOMI data with *in situ* monitoring stations has improved the precision of predicting PM_{10} concentrations. Machine learning approaches, especially ensemble learning algorithms, have demonstrated efficacy in managing the complexity of air quality data, scoring the highest accuracy during the heating season with RMSE of 4.52 μg/m³, MSE of 20.44 (μg/m³)², and MAE of 3.30 μg/m³ for validation and RMSE of 4.38 μg/m³, MSE of 19.21 (μg/m³)², and MAE of 3.19 μg/m³ for testing.

Assess the potential for AI to strengthen air pollution management, thereby contributing to public health and policy-making. The incorporation of GIS modelling allows the visualisation and contextualisation of PM_{10} impacts across various seasons, demonstrating pollution trends that indicate superior air quality during the heating season compared to the non-heating season, thereby assessing environmental management and decision-making processes.

References

Ayturan, A. Y., Ayturan, Z. C., & Altun, H. O. (2018). Air Pollution Modelling with Deep Learning: A Review. *Int. J. of Environmental Pollution & Environmental Modelling, 1*(3), 58–62. https://www.researchgate.net/publication/328927914

Bates, S., Hastie, T., & Tibshirani, R. (2022). *Cross-validation: what does it estimate and how well does it do it?* https://doi.org/10.1080/01621459.2023.2197686

Bbeiman, L. (1996). Bagging Predictors. *Machine Learning, 24*, 123–140. https://doi.org/10.1007/BF00058655

Breslow, L. A., & Aha, D. W. (1997). Simplifying decision trees: A survey. *The Knowledge Engineering Review*, *12*(01), 1–40. https://doi.org/10.1017/S0269888997000015

Chen, L., Han, B., Wang, X., Zhao, J., Yang, W., & Yang, Z. (2023). Machine Learning Methods in Weather and Climate Applications: A Survey. *Appl. Sci. 13*, 12019. https://doi.org/10.20944/preprints202309.1764.v2

Das, A., & Sahu, M. (2024). Leveraging Satellite Data for Predicting PM10 Concentration with Machine Learning Models: A Study in the Plains of North Bengal, India. *Aerosol and Air Quality Research*, *24*(12), 240066. https://doi.org/10.4209/AAQR.240066

European Environment Agency. (n.d.). *Home page.* https://www.eea.europa.eu/en

Ecosystem. (2025, April 23). *Copernicus Data Space Ecosystem – Europe's eyes on Earth*. Copernicus Data Space Ecosystem. https://dataspace.copernicus.eu/

Hu, K., Rahman, A., Bhrugubanda, H., & Sivaraman, V. (2017). HazeEst: Machine Learning Based Metropolitan Air Pollution Estimation From Fixed and Mobile Sensors. *IEEE Sensors Journal*, *17*(11), 3517–3525. https://doi.org/10.1109/JSEN.2017.2690975

Ketu, S. (2022). Spatial Air Quality Index and Air Pollutant Concentration prediction using Linear Regression based Recursive Feature Elimination with Random Forest Regression (RFERF): a case study in India. *Natural Hazards*, *114*(2), 2109–2138.
https://doi.org/10.1007/s11069-022-05463-z

Larkin, A., Geddes, J. A., Martin, R. V, Xiao, Q., Liu, Y., Marshall, J. D., Brauer, M., & Hystad, P. (2017). Global Land Use Regression Model for Nitrogen Dioxide Air Pollution. *Environmental Science & Technology*, *51*(12), 6957–6964. https://doi.org/10.1021/acs.est.7b01148

Li, J., & Heap, A. D. (2011). A review of comparative studies of spatial interpolation methods in environmental sciences: Performance and impact factors. *Ecological Informatics*, *6*(3), 228–241. https://doi.org/10.1016/j.ecoinf.2010.12.003

Lin, L., Liang, Y., Liu, L., Zhang, Y., Xie, D., Yin, F., & Ashraf, T. (2022). Estimating PM2.5 Concentrations Using the Machine Learning RF-XGBoost Model in Guanzhong Urban Agglomeration, China. *Remote Sensing*, *14*(20).
https://doi.org/10.3390/rs14205239

Lu, G. Y., & Wong, D. W. (2008). An adaptive inverse-distance weighting spatial interpolation technique. *Computers and Geosciences*, *34*(9), 1044–1055.
https://doi.org/10.1016/j.cageo.2007.07.010

Mamić, L., Gašparović, M., & Kaplan, G. (2023). Developing PM2.5 and PM10 prediction models on a national and regional scale using open-source remote sensing data. *Environmental Monitoring and Assessment*, *195*(644).
https://doi.org/10.1007/s10661-023-11212-x

Rodríguez-García, M. I., Carrasco-García, M. G., Ribeiro, M. da C. R., González-Enrique, J., Ruiz-Aguilar, J. J., & Turias, I. J. (2024). Air Pollution PM10 Forecasting Maps in the Maritime Area of the Bay of Algeciras (Spain). *Journal of Marine Science and Engineering*, *12*(3).
https://doi.org/10.3390/jmse12030397

Rybarczyk, Y., & Zalakeviciute, R. (2018). Machine Learning Approaches for Outdoor Air Quality Modelling: A Systematic Review. *Applied Sciences, 8*(12), 2570.
https://doi.org/10.3390/app8122570

Saradjian, M. R., & Akhoondzadeh, M. (2011). Thermal anomalies detection before strong earthquakes (M >6.0) using interquartile, wavelet and Kalman filter methods. *Hazards Earth Syst. Sci*, *11*, 1099–1108.
https://doi.org/10.5194/nhess-11-1099-2011

Shah, S. V., Gaikwad, S. V., & Vibhute, A. D. (2024). Air Quality Monitoring Using Sentinel-5p TROPOMI—A Case Study of Pune City. *SN Computer Science*, *5*(8).
https://doi.org/10.1007/s42979-024-03500-1

Shepard, D. (n.d.). A two-dimensional interpolation for irregularly-spaced data function. In *Proceedings of the 1968 ACM National Conference* (pp. 517–524). Association for Computing Machinery. https://doi.org/10.1145/800186.810616

Stanford CS Theory. (n.d.). Retrieved September 10, 2024, from https://theory.stanford.edu/main/index.shtml

Tredennick, A. T., Hooker, G., Ellner, S. P., & Adler, P. B. (2021). A practical guide to selecting models for exploration, inference, and prediction in ecology. *Ecology, 102*(6).
https://doi.org/10.1002/ecy.3336

DI ALGORITMŲ TAIKYMAS PROJEKTUOJAMOJO ERDVĖS LAIKO PM_{10} IR ORO TARŠAI MODELIUOTI

M. A. S. FAHIM,
J. SUŽIEDELYTĖ VISOCKIENĖ,
R. GRUBLIAUSKAS

Santrauka. Oro tarša yra viena iš rimčiausių šių dienų pasaulio problemų. Ji kelia didelį pavojų žmonių sveikatai ir aplinkai, skatina kvėpavimo takų ligas ir prisideda prie klimato kaitos. Oro teršalų identifikavimas ir oro kokybės valdymas yra svarbiausi veiksniai siekiant tikslesnio taršos prognozavimo ir modeliavimo. Tyrimo tikslas – įvertinti kietųjų dalelių, kurių skersmuo mažesnis nei 10 μm (PM_{10}), koncentracijas ir atlikti jų prognozavimą, remiantis Europos Sąjungos „Copernicus“ programos duomenimis. Naudojami Sentinel-5P palydovo troposferos stebėjimo instrumento „Tropomi“ duomenys, gauti iš stebėjimo stočių, išsidėsčiusių visoje Europoje. Sukurto modelio tikslumas vertinamas pagal vidutinę kvadratinę (RMSE) ir vidutinę absoliutinę (MAE) paklaidas. Tyrime taikomi mašininio mokymosi algoritmai, naudojant įvairius duomenų rinkinius. Pasitelkus geografinių informacinių sistemų (GIS) įrankius, atliktas erdvinis PM_{10} koncentracijų modeliavimas.

Reikšminiai žodžiai: GIS, kietosios dalelės, mašininis mokymasis, meteorologiniai duomenys, nuspėjamasis tikslumas, nuotolinis stebėjimas, Sentinel-5P TROPOMI.