Prediction of PM2.5 and PM10 Concentrations Using Machine and Deep Learning Techniques: Evidence from Tashkent, Uzbekistan

Sokhobiddin Akhatkulov[1, 2], Islom Yalgoshev[1], Izhar Uddin[1, 3], Abubakir Abdullayev[4], Shohkrukh Sariyev[1] and Rakhim Dusanov[1]

[1]*Department of Control Theory and Information Security, Samarkand State University, Universitet Blvd. 15, 140104 Samarkand, Uzbekistan*

[2]*Department of Exact Sciences, Samarkand Branch of Kimyo International University in Tashkent, Kh. Abdullaev Str. 63, 140143 Samarkand, Uzbekistan*

[3]*Department of Mathematics, Jamia Millia Islamia University, 110025 New Delhi, Delhi, India*

[4]*Department of Informatics, Samarkand State Pedagogical Institute, Spitamen Ave. 166, 140103 Samarkand, Uzbekistan*

akhatkulov@gmail.com, islomyalgoshev95@gmail.com, izharuddin1@jmi.ac.in, abdullayev73abubakir@gmail.com, rahimdosanov966@gmail.com, shokhrukhsariyev95@gmail.com

Keywords: PM2.5, PM10, Air Quality, Machine and Deep Learning Models.

Abstract: This study investigates several machine and deep learning models to predict the concentrations of PM2.5 and PM10 particles in the air of Tashkent city. These particulate matters are the most influential factors in Tashkent's high air quality index. The dataset utilized in this investigation comprises PM2.5 and PM10 concentration levels alongside meteorological indicators, including temperature, humidity, wind speed, and air pressure, gathered from 16 monitoring stations around Tashkent city and accessible at https://opendata.tashkent.uz/. The algorithms, Linear Regression (LR), Random Forest (RF), Support Vector Regression (SVR) and Gated Recurrent Unit (GRU) neural network were used to build predictive models on the concentrations of particles PM2.5 and PM10. The model performances were evaluated using metrics RMSE, MAE, MAPE, R^2 and compared. The results show that the GRU and RF models performed significantly better than LR and SVR with R^2 values of 0.98 and 0.97, respectively, compared to the value of 0.91 and 0.93 respectively, achieved by LR and SVR. The GRU model was evaluated as the most effective approach due to its ability to deeply explore dynamic relationships across time series.

1 INTRODUCTION

Air pollution is a major threat to the health of people and the environment all around the world. The World Health Organization claims that outdoor air pollution kills over 4.2 million people every year. Almost 98 percent of people throughout the world live in air that is worse than what the WHO says is safe [1]. When looking at air quality, PM2.5 and PM10 are very important. PM2.5 stands for particles that are less than 2.5 micrometers in diameter, and PM10 stands for particles that are up to 10 micrometers in size. These particles can get into the lungs and the circulatory system through the respiratory tract. This raises the risk of cardiovascular and respiratory disorders, including cancer and early death [2]. In recent years, Tashkent has also shown disturbingly bad signs of air pollution with particle matter. High road traffic density, emissions from heating and industrial businesses, and weather circumstances including temperature inversion, low wind speed, and dust storms during the dry season are some of the things that create this. In addition to constantly checking the air quality, it is important to be able to estimate PM2.5 and PM10 levels over the next few hours and days so that people can be warned in time and the right management actions can be made. ARIMA and linear regression are examples of traditional statistical models that can help predict air quality. However, these models typically don't work well because air pollution indicators are not always clear-cut and can be affected by many other things. Consequently, the utilization of machine learning and deep learning methodologies has gained significant prominence in recent years. Classical methodologies, including Random Forest, Gradient Boosting, and Support Vector Regression, alongside contemporary

neural network designs such as LSTM, GRU, and CNN-GRU, have been effectively utilized to predict PM2.5 and PM10 concentrations in numerous global cities [3], [12].

This study systematically evaluated LR, SVR, RF, and GRU for forecasting PM2.5 and PM10 in Tashkent. The accuracy of GRU was the highest: PM2.5: R2=0.98, MAE=2.14; PM10: R2=0.98, MAE=2.71. The accuracy of RF was next: PM2.5: R2=0.97; PM10: R2=0.96. Both did far better than the persistence baseline. GRU demonstrated the highest performance in the considered dataset and may serve as a promising basis for operational early-warning applications in Tashkent. In the future, we will look into CNN-GRU hybrids and expand the framework to include other cities in Central Asia.

- A carefully chosen open-source dataset for Tashkent has been put together. It includes hourly PM2.5, PM10 measurements from 16 stations and meteorological observations taken at the same time.
- A systematic lag and seasonality feature engineering pipeline is suggested, which captures daily and weekly patterns in how particulate matter moves.
- A strict historical comparison of standard ML approaches LR, SVR, RF and a deep learning model GRU is done with the same data.
- There are practical tips for setting up a real-time PM forecast system and early-warning criteria for Tashkent's air quality monitoring network.

Unlike the majority of published studies that either focus on single monitoring points or rely on globally aggregated datasets, the present work develops a region-specific, high-resolution, multi-station time-series forecasting framework tailored to Tashkent's urban environment. The pipeline integrates spatial aggregation across 16 distributed monitoring nodes with a lag- and seasonality-based feature engineering scheme that accounts for the characteristic meteorological patterns of Central Asia, including dust-storm episodes, temperature inversions, and distinct heating-season dynamics. To the authors' knowledge, no comparable end-to-end benchmark has previously been reported for this geography. Such region-specific monitoring and data-driven ecological evaluation frameworks align with recent advancements in localized geospatial monitoring and mathematical modeling platforms developed for infrastructure and environmental tracking in Central Asia [5], [6].

2 LITERATURE REVIEW

Two main directions can be seen in the international scientific literature on air pollution prediction: numerical models based on physicochemical processes, such as Euler and Lagrange dispersion models, and data-driven approaches based on statistics and machine learning [4]. For data-driven approaches, it was observed that initially models such as linear regression, ARIMA, and SARIMA were widely used. Such models are straightforward, fast, and readily interpretable, but they have shortcomings in reflecting complex nonlinear relationships. Later, various classical learning algorithms, including k-nearest neighbors, decision trees, and Random Forest, began to be widely used for estimating PM2.5 and PM10 levels in atmospheric air.

A study on the city of Delhi examined the outcomes of linear regression, decision tree, Random Forest, KNN, and Lasso regression models. It was found that the Random Forest and XGBoost models had the lowest error criteria, MAE and RMSE [3], [8], [9]. In recent years, approaches based on deep neural network models, including LSTM and GRU from the RNN family, and hybrid frameworks like CNN-LSTM and CNN-GRU, have demonstrated high accuracy in numerous studies. Ling and co-authors proposed a hybrid GRA-GRU model for predicting PM2.5 concentrations, and reported that it outperformed LSTM, simple GRU, and classical models [7], [10]-[13].

Shakya and co-authors apply the GRU-based encoder–decoder model for one-hour, eight-hour, and daily PM2.5 forecasts and emphasize that the GRU architecture is highly effective for modeling complex time series. When reviewing studies aimed at jointly predicting PM2.5 and PM10, accuracy was improved using sequential or overlapping LSTM and GRU methods, as well as through or optimized hyperparameters [14]. In a work published in 2025, the LSTM-GRU stack model was used to predict PM2.5 and PM10 in a formal ELV treatment zone using multiple outputs and was shown to outperform classical machine learning models [15]-[17]. Aggregation-based methods such as Random Forest and XGBoost provide stable and high accuracy. in Indian cities, Random Forest showed R^2 up to 0.93 for PM prediction [18]. SVR works well, especially for medium-sized data, if an appropriate kernel is chosen. RNN models such as GRU and LSTM can provide the best results for problems with strong temporal dependencies [19]-[22]. It has also been shown that combining deep learning and ensemble methods works well for predicting environmental emissions in

other fields. For example, Akhatkulov et al. used GRU and ensemble ML models to predict vehicle CO2 emissions and got good results with similar feature engineering techniques [23]. On this basis, it was found appropriate to select Linear Regression, Random Forest, SVR, and GRU models for the conditions of Tashkent in this study.

3 METHODOLOGY

In this section, we briefly consider the machine and deep learning algorithms which we use to build our predictive models. For more detailed explanations of these algorithms see [24].

Multivariate linear regression assumes a linear regression is based on the assumption of a linear association between the dependent variable y PM10 or PM2.5 and the explanatory factors x_1 and $x_n$concentration, temperature, humidity, etc. in previous periods:

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n + \varepsilon,$$

where β is the regression coefficients, and ε represents the random error. The given indicators are estimated based on the least squares method. Although Support Vector Machine Regression was originally designed for classification, the Support Vector Regression option can also be used to solve continuous prediction tasks. It is reflected that SVR seeks to keep the objective function within the insensitive band ε.

$$\min_{w,b,\xi_j,\xi_j^*} \frac{1}{2}\|W\|^2 + C\sum_{j=1}^{N}(\xi_j + \xi_j^*)$$
$$\begin{cases} y_j - (W \cdot x_j + b) \le \varepsilon + \xi_j, \\ (W \cdot x_j + b) - y_j \le \varepsilon + \xi_j^*, \\ \xi_j, \xi_j^* \ge 0. \end{cases}$$

Here C is the penalty coefficient, ε is the allowable error, W and b are the model parameters. The kernel function $K(x_i, x_j)$ can also be used to transform data into a higher-dimensional space and model nonlinear relationships.

Random Forest model. Random Forest is an ensemble method based on the bagging principle and uses the average result of many decision trees, namely $T_1(x)$,$T_M(x)$:

$$\hat{y} = \frac{1}{M}\sum_{m=1}^{M} T_m(x).$$

Each tree is constructed using randomly resampled training data, while a randomly selected group of input features is evaluated at every split. This reduces the correlation between trees and thereby enhances the general accuracy of the ensemble. Random Forest captures nonlinear relationships well and is robust to outliers, allowing for feature importance estimation [25]-[27]. However, it faces problems in interpreting this multi-tree ensemble.

GRU neural network. Gated recurrent unit is an RNN architecture designed for time series and sequences, and is simpler than LSTM. At each time point, the GRU cell is described through the set of equations given below [28]:

$$\begin{aligned} z_t &= \sigma(W_z x_z + U_z d_{t-1} + b_z), \\ r_t &= \sigma(W_r x_t + U_r d_{t-1} + b_r), \\ \tilde{d}_t &= tanh \quad (W_d x_t + U_h(r_t \odot d_{t-1}) + b_h), \\ d_t &= (1 - z_t) \odot d_{t-1} + z_t \odot \tilde{d}_t. \end{aligned}$$

Here, z_t is the update gate, r_t is the reset gate, d_t is the hidden state, σ is the sigmoid, the hyperbolic tangent function, and $\odot$ represents the component-wise multiplication. The GRU differs from the LSTM in that it has fewer parameters and, in many cases, differs in training speed and generalization ability.

4 DATA AND RESEARCH DESIGN

The dataset utilized in this research was sourced from automated air quality and meteorological monitoring stations situated across Tashkent city, accessible at https://opendata.tashkent.uz/. The dataset was gathered from 16 monitoring stations located in various districts of Tashkent, which mainly collect data on the air quality index, air contaminants, and weather conditions. The dataset goes from November 7, 2023, to January 5, 2025, which is around 14 months. Every hour, data were captured, resulting in a total of 164,505 observations across all 16 sites. The final cleaned dataset had 158,720 samples that may be used after pre-processing. The following characteristics were found to be present for each observation:

Date and time; PM2.5 (μg/m³); PM10 (μg/m³); air temperature (°C); relative humidity (%); wind speed (m/s); atmospheric pressure (hPa).

4.1 Pre-Processing

By merging hourly data from 16 monitoring sites, each record was given a station ID and geographic coordinates. Rows that had more than 30% of their values missing were taken out. For short missing portions of up to 3 hours in a row, linear interpolation was used to fill them in. For larger gaps, the station-specific average for the same hour of day and month was used. Values that were lower than Q1 − 3×IQR or higher than Q3 + 3×IQR were considered outliers and were replaced with a 24-hour rolling median. For PM2.5, PM10, temperature, humidity, and wind speed, lagged predictors were made for 1, 2, 3, 6, 12, and 24 hours before the time of the forecast. The seasonal variables were the hour of the day, the day of the week, the month, and a binary indicator that showed whether it was day or night. Finally, we used Min-Max scaling to normalize all of the numerical variables, but solely on the training data [29]–[31].

4.2 Train–Test Split and Cross-Validation

A precise chronological split was employed to divide the data into sets for training and testing. The training set was made up of 80% of the data from November 7, 2023 to October 31, 2024. The test set was made up of all the data from November 1, 2024 to January 5, 2025. The overall architecture of the proposed forecasting and early-warning system is summarised in Figure 1. Sensor readings from 16 monitoring stations feed a centralised data-collection layer, where incoming streams are timestamped, station-labelled, and stored in a raw repository.

The preprocessing module then applies the cleaning and feature-engineering steps described in Section 4.1. The processed feature vectors serve as input to the four candidate models LR, SVR, RF, GRU, each generating one-hour-ahead PM2.5 and PM10 forecasts. A threshold-based decision layer compares the predicted concentrations against WHO guideline values PM2.5 > 15 µg/m³; PM10 > 45 µg/m³ for 24-h mean and issues colour-coded alerts - green, yellow, or red - to the operator dashboard and, where applicable, to municipal notification channels. During feature engineering, imputation, normalization, model training, or hyperparameter tuning, no information from the test period was used to keep data from leaking. We merely fit Min-Max scaling to the training data, and then we used it on the validation and test subsets without recalibrating it [29]-[31]. We only used past observations to make all of the lagged variables; we didn't use any future timestamps. Similarly, station-level information employed for imputation were derived exclusively from the training segment. We used a five-fold expanding-window cross-validation approach to tune the hyperparameters in the training set. During each fold, the training time was gradually lengthened, and the model's performance was tested during the validation period that followed.

The resulting models were also evaluated on three different, non-overlapping time periods November 2024, December 2024, and January 1-5, 2025 to see if they were stable over time. We chose these time periods to show how pollution and the seasons change. The GRU and RF models consistently had the lowest MAE and RMSE values over all three test periods. This shows that the results are stable and reliable. In Linear Regression, you can try one of the Simple, Ridge, and Lasso options. Finding RBF kernel parameters and gamma in SVR with grid search. In random forest, parameters such as the total trees, maximum tree depth, and smallest leaf node size are optimized. In GRU, the hidden state size, the total layers, learning pace, batch size, total training cycles, and dropout coefficient are optimized using the Adam optimizer. Assessment criteria the following criteria are considered for PM2.5 and PM10 for each model: MAE, RMSE, MAPE and R^2.

These criteria have traditionally been used in air pollution forecasts and are also widely used in comparing machine learning models. Below is a summary of the final hyperparameter settings chosen by time-series cross-validation. The RBF kernel was used for SVR, with a regularization value of $C = 10$, a kernel coefficient of $\gamma = 0.1$ and an epsilon of $\varepsilon = 0.1$. For Random Forest, the values are n_estimators=200, max_depth=20 and min_samples_leaf = 2. The GRU model has two recurrent layers, each with 64 hidden units, a dropout rate of 0.2, a look-back sequence length of 24 hours, a learning rate of 0.001 Adam optimizer, a batch size of 128, and training for 100 epochs with early stopping. It was possible to estimate linear regression using conventional least squares, hence there was no need to tune any hyperparameters.

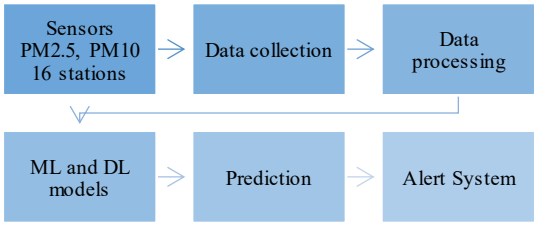

Figure 1: Proposed architecture of the PM2.5/PM10 forecasting and alert system.

5 RESULTS

A naive persistence model was added as a baseline reference. This model says that the current concentration is the same as the value from the previous time step:

$$PM(t) = PM(t-1).$$

For PM2.5, this baseline gave us MAE = 9.63, RMSE = 12.14, MAPE = 15.87%, and $R^2 = 0.81$. The MAE for PM10 was 12.48, the RMSE was 15.36, the MAPE was 18.92%, and the R^2 was 0.77.

All trained models greatly exceed this baseline, validating that the implemented features and model topologies yield considerable predictive enhancement beyond mere one-step persistence. The research evaluated the performance of linear regression, SVR, RF, and GRU models to predict levels of PM2.5 and PM10 in the atmospheric air of Tashkent. The MAE, MSE, RMSE, MAPE, and R^2 evaluation criteria were calculated separately for each model result. Figure 2 shows the comparison of the visual results obtained by the machine learning models, comparing the MAE and RMSE metrics related to PM2.5 and PM10.

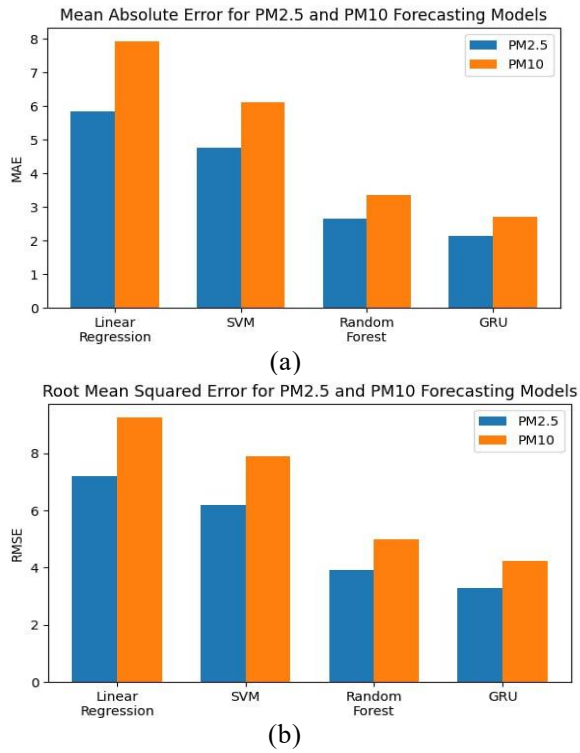

Figure 2: Comparison of MAE and RMSE for PM2.5 and PM10 forecasting models: a) PM2.5; b) PM10.

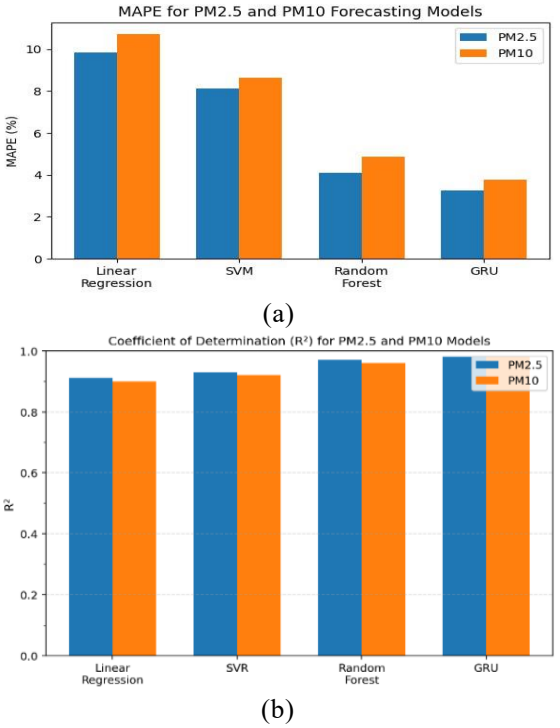

Figure 3: Comparison of MAPE and R^2 for PM2.5 and PM10 forecasting models: a) PM2.5; b) PM10.

The findings of this study are presented in Figure 3, where a graphical comparison of the MAPE and R^2 indicators is shown, and these outcomes indicate that the GRU model achieved strong performance. According to the results, the accuracy of the GRU and Random Forest models was the highest.

The GRU model seems to work better since PM2.5 and PM10 levels in Tashkent change over time. This is because the levels of these pollutants follow clear patterns that change with the seasons and weather. These instances happen in the winter when there is no movement and in the summer months when there is dust. The GRU architecture is built to pick up on longer-term relationships while downplaying the effects of short-term changes.

This could be why it works better for pollution predictions. The RF model also performed well, probably because it used lagged and seasonal variables to capture a lot of the data's time structure. When pollution levels were high during the test period, the prediction error went up less for GRU than for RF and LR. This suggests that GRU was more stable during extreme events. From a practical standpoint, the reported PM2.5 inaccuracy was rather minor when compared to standard air-quality reference values, and the model predictions were

generally consistent enough to maintain the same air-quality interpretation in the majority of test scenarios. The results of the real-time dataset from Tashkent, Uzbekistan, are presented in Table 1 below, which provides the results of forecasting the PM2.5 levels in atmospheric air. For the linear regression model results, MAE = 5.84, RMSE = 7.22, and R^2 = 0.91, while the SVR model yielded MAE = 4.76, RMSE = 6.20, and R^2 = 0.93, which showed a better performance than linear regression.

Table 1: Model results in PM2.5 forecasting.

Model	MAE	RMSE	MAPE	R^2
LR	5.84	7.22	9.85	0.91
SVR	4.76	6.20	8.12	0.93
RF	2.65	3.91	4.10	0.97
GRU	2.14	3.29	3.25	0.98
Persistence	9.63	12.14	15.87	0.81

In the next step, the results obtained by machine learning models on this dataset for PM10 are presented in Table 2. For the linear regression model, MAE = 7.92, RMSE = 9.26, and R^2 = 0.90 were recorded, which did not show high accuracy, although this model also served as a baseline approach for PM10. The SVR model performed slightly better than the linear model, with MAE = 6.11, RMSE = 7.89, and R^2 = 0.92.

Table 2: Model results in PM10 forecasting.

Model	MAE	RMSE	MAPE	R^2
LR	7.92	9.26	10.72	0.90
SVR	6.11	7.89	8.64	0.92
RF	3.34	4.99	4.85	0.96
GRU	2.71	4.24	3.78	0.98
Persistence	12.48	15.36	18.92	0.77

Overall, the results show that the GRU model, specially adapted for time series, showed the highest accuracy in predicting PM2.5 and PM10 levels in the atmospheric air of Tashkent, while Random Forest is a powerful alternative that is relatively simple in terms of computation and easy to interpret, while also having high accuracy.

Based on the results of the tables and graphs, we can expect and analyze the following trends. For linear regression, we present the following. The model is simple and fast, but it was observed that it does not well capture the nonlinear relationships between PM2.5 and PM10 dynamics and meteorological factors. It was estimated that the MAE and RMSE were higher than other models, but the R^2 was more likely to be lower.

The following was revealed in SVR. Thanks to the RBF kernel, it can model some degree of nonlinearity. If hyperparameters are chosen correctly, they can provide much better results than linear regression but can be computationally slow on large amounts of data.

The following findings were made for random forest. It is a multi-tree ensemble that does a good job of capturing complicated connections and is resistant to outliers. This makes it useful for predictive modeling and lets it work with a wider range of datasets. It frequently gives the greatest or almost the best results compared to GRU when it comes to MAE, RMSE, and R^2=0.97.

The GRU model had the smallest inaccuracy in its predictions of all the models that were tested. It got MAE = 2.14, RMSE = 3.29, MAPE = 3.25%, and R^2 = 0.98 for PM2.5; and MAE = 2.71, RMSE = 4.24, MAPE = 3.78%, and R^2 = 0.98 for PM10. These results show that the GRU architecture can accurately represent long-term temporal dependencies, daily patterns, and peak pollution events like dust storms and thermal inversions, whereas linear and tree-based models only partially do.

Despite the encouraging performance metrics, several limitations of the present study should be acknowledged. The dataset is confined to Tashkent city, and the trained models have not been evaluated on measurements from other urban or semi-arid environments; hence, the extent to which the proposed framework generalises beyond this geography remains an open question. Furthermore, the predictor set does not incorporate explicit proxies for key emission sources such as industrial stack discharges, traffic volume indices, or construction activity, all of which are known contributors to particulate loading in Tashkent's urban atmosphere. Omitting these covariates may introduce systematic bias during periods when anthropogenic source intensity deviates substantially from its climatological mean. Finally, the early-warning architecture outlined in this work remains at the design stage; real-time deployment, integration with municipal notification infrastructure, and prospective validation against independent sensor streams have not yet been carried out and constitute priority directions for subsequent research.

6 CONCLUSIONS

This study assessed the relative predictive capacity of four modelling approaches - linear regression, SVR, random forest, and GRU - applied to hourly PM2.5

and PM10 data collected across 16 monitoring stations in Tashkent city. Among the evaluated configurations, the GRU architecture demonstrated the highest forecasting accuracy on the considered dataset and under the specific meteorological and urban conditions of Tashkent, attaining MAE = 2.14, RMSE = 3.29, MAPE = 3.25% and R^2 = 0.98 for PM2.5, and comparable performance for PM10. The next most efficient solution was RF, which yielded MAE = 2.65, RMSE = 3.91, MAPE = 4.10, and R^2 = 0.97. The remaining two models had higher errors: SVR yielded MAE = 4.76, RMSE = 6.20, MAPE = 8.12, and R^2 = 0.93, while LR showed somewhat higher errors, with MAE = 5.84, RMSE = 7.22, and R^2 = 0.91. Based on the data obtained, GRU seems to be an appropriate choice for the operationalization of PM2.5 and PM10 forecasting in Tashkent, due to its ability to model long-range temporal dependencies and non-linear pollution dynamics. Random Forest provides a computationally efficient and more interpretable option with only a slight decrease in accuracy, which might be beneficial in resource-limited deployment scenarios. The applicability of these findings to other cities or climatic zones, however, requires further empirical investigation.

REFERENCES

[1] World Health Organization, "Ambient (outdoor) air quality and health," WHO Fact Sheet, 2024, [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/ambient-%28outdoor%29-air-quality-and-health.

[2] European Environment Agency, "Harm to human health from air pollution in Europe," EEA Report, 2024, [Online]. Available: https://www.eea.europa.eu/en/analysis/publications/harm-to-human-health-from-air-pollution-2024.

[3] R. P. Kumar, A. Prakash, R. Singh, et al., "Machine learning-based prediction of hazards fine PM2.5 concentrations: A case study of Delhi, India," Discovery Geoscience, vol. 2, Art. no. 34, 2024, [Online]. Available: https://doi.org/10.1007/s44288-024-00043-z.

[4] S. Mampitiya, et al., "Exploring PM2.5 and PM10 ML Forecasting Models: A Comparative Study in the UAE," Scientific Reports, vol. 15, Art. no. 9536, Mar. 2025, [Online]. Available: https://doi.org/10.1038/s41598-025-94013-1.

[5] M. Shukurova, M. Talipov, K. Ruziev, and K. Jurayeva, "Advanced geospatial monitoring of oil and gas infrastructure via satellite data," Mathematical Models in Engineering, vol. 12, no. 2, pp. 190-201, Jun. 2026, [Online]. Available: https://doi.org/10.21595/mme.2026.25328.

[6] R. Aliev, M. Aliev, and G. Talipova, "Mathematical model and algorithm for determining the shunt zone by train," AIP Conference Proceedings, vol. 3447, no. 1, Art. no. 020004, May 2026, [Online]. Available: https://doi.org/10.1063/12.0044000.

[7] A. Mampitiya, N. Rathnayake, Y. S. Leon, P. Hoshino, and U. Rathnayake, "Machine Learning Approaches for Predicting the Air Quality Index," Environmental Pollution, 2023, [Online]. Available: https://doi.org/10.1016/j.envpol.2023.122293.

[8] Q. Di, Y. Wang, et al., "An ensemble-based model of PM2.5 concentration across the contiguous United States with high spatiotemporal resolution," Environment International, vol. 141, Art. no. 105726, 2020, [Online]. Available: https://doi.org/10.1016/j.envint.2020.105726.

[9] A. Masood and K. Ahmad, "Data-Driven Predictive Modeling of PM2.5 Concentrations Using Machine Learning and Deep Learning Techniques: A Case Study of Delhi, India," Environmental Monitoring and Assessment, vol. 195, Art. no. 60, 2022, [Online]. Available: https://doi.org/10.1007/s10661-022-10603-w.

[10] L. Qing, "PM2.5 Concentration Prediction Using GRA-GRU Network," Sustainability, vol. 15, no. 3, Art. no. 1973, 2023, [Online]. Available: https://doi.org/10.3390/su15031973.

[11] L. Dai, C. Zhang, and M. Lei, "Dynamic forecasting model of short-term PM2.5 concentration based on machine learning," Journal of Computer Applications, vol. 37, pp. 3057-3063, 2017.

[12] X. Wu, J. Zhu, and Q. Wen, "Short-Term Prediction of PM2.5 Concentration by Hybrid Neural Network Based on Sequence Decomposition," PLOS ONE, vol. 19, no. 5, Art. no. e0299603, May 2024, [Online]. Available: https://doi.org/10.1371/journal.pone.0299603.

[13] S. Zhou, W. Wang, L. Zhu, Q. Qiao, and Y. Kang, "Deep-Learning Architecture for PM2.5 Concentration Prediction: A Review," Environmental Science and Ecotechnology, vol. 21, Art. no. 100400, 2024, [Online]. Available: https://doi.org/10.1016/j.ese.2024.100400.

[14] A. Shakya, M. S. Gohain, and R. Singh, "A Systematic Study on PM2.5 and PM10 Concentration Prediction in Air Pollution Using Machine Learning and Deep Learning Model," Environmental Pollution, 2025, [Online]. Available: https://doi.org/10.1016/j.envpol.2025.122293.

[15] H. Alrashidi, F. N. Sibai, A. Abonamah, M. Alrashidi, and A. Alsaber, "PM2.5: Air Quality Index Prediction Using Machine Learning: Evidence from Kuwait's Air Quality Monitoring Stations," Sustainability, vol. 17, no. 20, Art. no. 9136, 2025, [Online]. Available: https://doi.org/10.3390/su17209136.

[16] C. Chen, et al., "A deep learning approach to identify smoke plumes in satellite imagery in near-real time for health risk communication," Environmental Health, vol. 20, Art. no. 11, 2021, [Online]. Available: https://doi.org/10.1186/s12940-021-00722-z.

[17] Z. Gao, K. Do, Z. Li, X. Jiang, K. J. Maji, C. E. Ivey, and A. G. Russell, "Predicting PM2.5 levels and exceedance days using machine learning methods," Atmospheric Environment, vol. 321, Art. no. 120293, 2024, [Online]. Available: https://doi.org/10.1016/j.atmosenv.2024.120293.

[18] K. Kumar and D. B. Pande, "Air Pollution Prediction with Machine Learning: A Case Study of Indian Cities," International Journal of Environmental Science and Technology, vol. 20, pp. 5333-5348, 2022, [Online]. Available: https://doi.org/10.1007/s13762-022-04241-5.

[19] V. N. Vapnik, The Nature of Statistical Learning Theory. New York, NY, USA: Springer, 1995.

[20] S. Akhatkulov, I. Yalgoshev, and J. Haydarov, "Different Warmup and Annealing Strategies for ANN Models to Predict Air Quality Index," in Proc. 2025 International Russian Automation Conference (RusAutoCon), Sochi, Russian Federation, pp. 491-496, 2025, [Online]. Available: https://doi.org/10.1109/RusAutoCon65989.2025.11177428.

[21] S. Liu, B. Li, and G. Hu, "Prediction of PM2.5 concentration based on random forest algorithm and meteorological data," Atmosphere, vol. 13, no. 4, Art. no. 521, 2022, [Online]. Available: https://doi.org/10.3390/atmos13040521.

[22] Y. Wu and Q. Wang, "LightGBM Based Optiver Realized Volatility Prediction," in Proc. 2021 IEEE International Conference on Computer Science, Artificial Intelligence and Electronic Engineering (CSAIEE), SC, USA, pp. 227-230, 2021, [Online]. Available: https://doi.org/10.1109/CSAIEE54046.2021.9543438.

[23] S. Akhatkulov, I. Yalgoshev, and Z. Urinboyev, "Vehicle CO2 Emission Prediction Using Deep Learning and Ensemble Machine Learning Methods," in Proc. 2025 International Russian Automation Conference (RusAutoCon), Sochi, Russian Federation, pp. 819-824, 2025, [Online]. Available: https://doi.org/10.1109/RusAutoCon65989.2025.11177377.

[24] P. Mahajan, S. Uddin, F. Hajati, and M. A. Moni, "Ensemble Learning for Disease Prediction: A Review," Healthcare, vol. 11, no. 12, Art. no. 1808, Jun. 2023, [Online]. Available: https://doi.org/10.3390/healthcare11121808.

[25] L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5-32, 2001, [Online]. Available: https://doi.org/10.1023/A:1010933404324.

[26] M. H. Abdulameer and M. Z. Abdullah, "Datasets Classification Using Deep Learning and Machine Learning Classification Algorithms," AIP Conference Proceedings, vol. 2591, no. 1, Art. no. 030032, 2023, [Online]. Available: https://doi.org/10.1063/5.0120454.

[27] J. Zhang, B. Li, and X. Chen, "PM2.5 concentration prediction with GRU and meteorological features in Beijing," Atmospheric Pollution Research, vol. 14, no. 6, Art. no. 101765, 2023, [Online]. Available: https://doi.org/10.1016/j.apr.2023.101765.

[28] K. Cho, B. van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning Phrase Representations Using RNN Encoder-Decoder for Statistical Machine Translation," arXiv preprint arXiv:1406.1078, 2014, [Online]. Available: https://doi.org/10.48550/arXiv.1406.1078.

[29] H. Bui, et al., "PM2.5 prediction using Random Forest algorithm: A case study of Hanoi, Vietnam," Science of the Total Environment, vol. 818, Art. no. 151791, 2022, [Online]. Available: https://doi.org/10.1016/j.scitotenv.2021.151791.

[30] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), New York, NY, USA, pp. 785-794, 2016.

[31] L. Gosink, et al., "Characterizing and Visualizing Predictive Uncertainty in Numerical Ensembles Through Bayesian Model Averaging," IEEE Transactions on Visualization and Computer Graphics, vol. 19, no. 12, pp. 2703-2712, Dec. 2013, [Online]. Available: https://doi.org/10.1109/TVCG.2013.138.