SAKARYA UNIVERSITY JOURNAL OF COMPUTER AND INFORMATION SCIENCES
http://saucis.sakarya.edu.tr/
e-ISSN: 2636-8129
Publisher: Sakarya University
Vol. 8, No. 1, 89-111, 2025
DOI: 10.35377/saucis...1560377

RESEARCH ARTICLE

Comparative Analysis of Data Visualization and Deep Learning Models in Air Quality Forecasting

Damla Mengus[1*], Bihter Daş[2]

[1]Marmara University, Computer Engineering, İstanbul, Türkiye, ror.org/02kswqa67
[2]Fırat University, Elazığ, Türkiye, ror.org/05teb7b63

Corresponding author:
Damla Mengus
Computer Engineering,
Marmara University, İstanbul, Türkiye
damla.mengus@marmara.edu.tr

Article History:
Received: 03.10.2024
Revised: 13.12.2024
Accepted: 09.01.2025
Published Online: 27.03.2025

ABSTRACT
This study utilizes air pollution data from the Continuous Monitoring Center of the Ministry of Environment, Urbanization, and Climate Change in Turkey to predict various pollutants using three advanced deep learning approaches: LSTM (Long Short-Term Memory), CNN (Convolutional Neural Network), and RNN (Recurrent Neural Network). Missing data in the dataset were imputed using the K-Nearest Neighbor (K-NN) algorithm to ensure data completeness. Furthermore, a data fusion technique was applied to integrate multiple pollutant enhancing the richness and reliability of the input features for modeling. The increasing air pollution issue, driven by factors such as population growth, urbanization, and industrial development, is a major environmental concern. The study evaluates these models to estimate pollutant concentrations and selects the most accurate, RNN, for forecasting air pollution over the next three years. Each prediction was assessed using performance metrics such as MAE, RMSE, and R^2 to ensure robust model evaluation. Visualization of the data and forecast results was achieved through methods like Box Plots, Violin Plots, and Point Scatter Graphs, making air quality information more accessible to general audiences. In terms of model performance, CNN achieved an R^2 of 0.88 for PM10 and 0.93 for SO2, while LSTM demonstrated an R^2 of 0.94 for PM10 and 0.95 for SO2. However, RNN emerged as the most accurate model, achieving an R^2 of 0.97 for both PM10 and SO2 forecasts. This model allows for forecasts of pollutant levels over a three-year period. The findings indicate that predictive modeling, combined with data fusion and visualization techniques, could significantly contribute to mitigating future uncertainties and enhance the comprehension of air quality patterns for non-expert audiences.

Keywords: Data prediction, CNN, RNN, LSTM, Data visualization

1. Introduction

Air pollution occurs when harmful substances are present in the atmosphere at levels that can negatively impact human health and the environment's equilibrium. Factors such as increasing population, rapid urbanization, and accelerated industrial growth contribute significantly to the degradation of air quality [1-3]. These trends have led to a rise in fossil fuel consumption, increased vehicular traffic, and expanded industrial activities. Major pollutants that contribute to air pollution include gases such as carbon monoxide (CO), nitrogen dioxide (NO2), sulfur dioxide (SO2), ozone (O3), and particulate matter (PM). Additionally, nitrogen oxides (NOx), particularly the interaction of nitric oxide (NO) and nitrogen dioxide (NO2), are central to atmospheric chemical reactions that further reduce air quality [4-6]. Particulate matter, especially PM10 and PM2.5, is among the most concerning pollutants due to its severe impact on human health. These particles are so small that they can bypass respiratory defenses, penetrating deep into the lungs, where prolonged exposure can lead to severe health issues, including cancer. When inhaled, SO2 is another dangerous pollutant that irritates the respiratory system and poses a considerable health risk [7-9]. Air quality is typically assessed using an index that reflects the level of pollution based on meteorological conditions. However, the air quality index may not always capture the true impact of air pollution due to measurement inaccuracies or insufficient sensor data, which can result in delayed responses. Although the air quality index is widely used to evaluate the harmful effects of pollution on public health, the sensitivity and precision of the measurements need improvement. Moreover, current measurements are often performed daily and cannot provide predictive insights, making it difficult to implement timely preventive measures. As the measured air quality index rises, it indicates worsening air conditions with increasing risks to human health. The index categorizes air quality from 0-50 as "good," 51-100 as "moderate," 101-150 as "sensitive," 151-200 as "unhealthy," 201-300 as "very unhealthy," and above 300 as "hazardous." Once the index surpasses 151, the likelihood of health problems increases, and outdoor activity becomes risky [10,11]. Figure 1 shows the air quality index.

Cite as: D. Mengus and B. Das, "Comparative analysis of data visualization and deep learning models in air quality forecasting," *Sakarya University Journal of Computer and Information Sciences*, vol. 8 no. 1, pp. 89-111, 2025, doi: 10.35377/saucis...1560377

In this study, missing data is systematically filled to address the issues of sensitivity and accuracy in air quality measurements, and comparative estimations are performed using three distinct deep learning approaches: RNN, CNN, and LSTM. The analysis identifies the model that provides the most accurate predictions, and this model is then used to forecast air quality over the next three years. The results are presented through visualizations to enhance clarity. One of the gaps in the current literature is the lack of comparative studies using these methods and the absence of hybrid models. This study seeks to bridge that gap by offering more visualization-based insights. The proposed solution provides a viable method to overcome existing air quality measurement limitations and generate future forecasts. Air quality data is collected from continuous monitoring stations (CMS) at 39 locations across Istanbul, making it highly effective for tracking air quality trends throughout the city and generating data-driven predictions. The location of Başakşehir district is depicted in Figure 2.

Figure 1. Air Quality Index

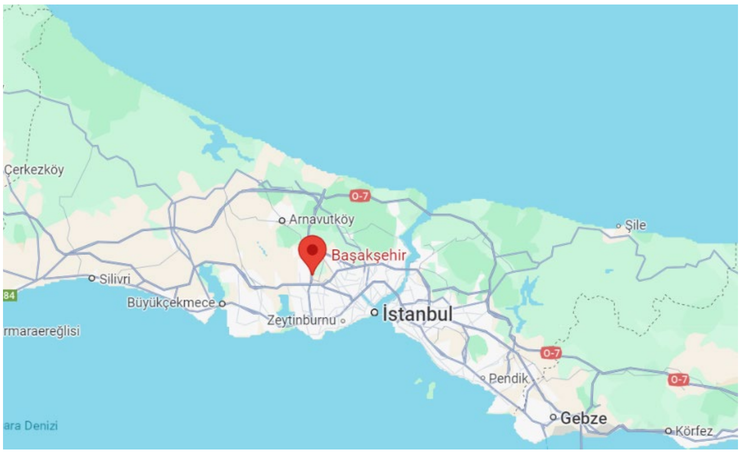

Figure 2. Location of Başakşehir District

1.1. Main Contributions

The key contributions of this study are as follows:

- LSTM, RNN, and CNN methods were used to predict air quality data, and these three methods were analyzed comparatively. As a result of the comparison, the method with the highest accuracy was determined, and air quality predictions for the next 3 years were made using this method. This study provides a comprehensive approach to visualizing large data sets and comparing the performances of different deep learning models, making significant contributions to environmental data analysis and prediction processes.
- Future air quality prediction will be made with RNN, which was determined as the best method. Thus, it is aimed at obtaining longer-term and more accurate predictions.
- By using Box Plots, Violin Plots, and Point Scatter Plots as visualization methods, non-experts helped interpret large data. This study aims to increase the effectiveness of data visualization and prediction methods. Thus, even non-expert users can interpret complex data sets.

1.2. Related Studies

Many methodologies and techniques have been applied in the literature to predict and visualize air quality. Deep learning and machine learning methods are especially prominent in air quality prediction. These methods have great potential in predicting the concentrations of various air pollutants and monitoring air quality. Below, a summary of important studies in this field is presented. In the article written by C.-J. Huang et al. aim to develop a model that combines deep learning methods LSTM and CNN networks to predict PM2.5 particulate matter. The model is applied to monitor and predict air quality in smart cities. It is a valuable study highlighting the effectiveness of hybrid deep learning models in air quality prediction for smart cities [12]. Similarly, R. Janarthanan et al. used deep learning approaches to predict the air quality index in a major metropolitan area. Their work highlights how deep learning can improve air quality forecasting in large urban settings [13]. S. Masmoudi et al. propose a machine learning framework that includes multi-objective regression and feature selection methods to predict the concentration of multiple air pollutants. This paper is an exemplary study that analyzes the application of multi-objective regression techniques in air pollution prediction [14]. Another study uses an attention mechanism model to predict air pollutant levels in the Yangtze River Delta during frequent heatwaves, improving accuracy in extreme weather conditions. The study provides a valuable approach by integrating an attention mechanism, offering improved air pollutant predictions during heatwaves, which could enhance environmental monitoring in climate-sensitive regions [15]. Another paper proposes a novel hybrid optimization model that considers other air pollutants and meteorological conditions to predict PM2.5 concentrations. It presents an approach highlighting hybrid models' importance in air pollution prediction [16]. P. A. Rani et al. developed a novel artificial intelligence algorithm that analyzes the levels of air pollutants to predict air quality in Tamil Nadu. This study investigates the innovative use of artificial intelligence techniques in air quality prediction [17]. Q. Wu and H. Lin have developed a new, highly efficient hybrid model for predicting the daily air quality index, incorporating air pollution factors. factors into account. This study is important research examining the effectiveness of hybrid model strategies in air quality prediction [18]. B. Zhang et al. predict the dispersion trends of air pollutants in specific seasons using deep learning. They examine the example of Northern China. They investigate the potential of deep learning applications in targeted seasonal analysis [19]. In another study, J. Luo and Y. Gong consider the prediction of air pollutants using a combination of ARIMA, Whale Optimization Algorithm (WOA), and LSTM models. It is aimed to increase the accuracy of air pollution prediction by integrating multiple algorithms [20]. G. I. Drewil and R. J. Al-Bahadili perform air pollution prediction using a combination of LSTM deep learning and metaheuristic algorithms. It is innovative research examining the integration of various algorithms in air pollution prediction [21]. Another study analyzes the thermal comfort indexes specific to the summer months in Istanbul and evaluates the effects of environmental variables. A comprehensive study has examined the effects of local thermal comfort and climate change [22]. In their paper, A. Kshirsagar and M. Shah deeply analyze using neural networks, regression, and hybrid models in air quality prediction. They present a comprehensive review of air quality prediction by comparing different model structures [23]. Another study reviews the existing literature on deep learning techniques for air pollutant concentration prediction and evaluates the application areas of these techniques. It has been a comprehensive review covering the advanced applications of deep learning techniques in air pollution prediction [24.] T. D. Akinosho et al. This paper uses a deep learning-based multi-objective regression model for traffic-related air pollution prediction. The model aims to improve its performance by simultaneously making predictions for multiple pollutants. This study demonstrates the potential of multi-objective deep learning models in traffic-related air pollution prediction [25]. M. Yılmaz et al. Their study detected the characteristics of heat waves in Istanbul and performed a regional analysis. The study provides a comprehensive analysis to understand the effects of climate change on cities. The study has been important research contributing to investigating urban climate change effects at the local level [26]. J. González-Pardo et al. This paper uses data mining models to predict changes in air pollutant levels in urban traffic areas in Spain during COVID-19 lockdown measures. It provides a valuable case study analyzing the changes in air pollution levels during the pandemic [27]. X. Shi et al. This paper evaluates the potential and additional benefits of reducing emissions of CO2 and other air pollutants from mobile sources in Shanghai. It is a case study investigating emission reduction strategies' environmental and economic benefits [28]. Another paper uses machine learning methods to predict air quality parameters and determine their spatial distribution. It emphasizes the importance of machine learning applications for spatial analysis in air quality prediction [29]. S. Ünaldi and N. Yalçin present a case study using machine learning methods for air pollution prediction in Başakşehir, Istanbul. It is a study examining the applicability of machine learning in local air quality prediction [30]. In the article, P. Aksak et al. examine the urban heat island effects and related climate parameters in Istanbul using remote sensing techniques. The study demonstrates the use of remote sensing data to understand the effects of the urban heat island phenomenon [31].

Y.-C. Lin et al.'s study uses Bayesian networks and deep learning models to evaluate the effects of meteorological and traffic factors on air pollutants. It analyzes using Bayesian networks and deep learning methods to model the factors affecting air quality [32]. I. H. Fong et al.'s study uses transfer learning and recurrent neural networks (RNN) to predict the concentration levels of air pollutants. It has been valuable research examining the role of transfer learning and RNN models in air pollution prediction [33]. Another study explores using a neural transfer learning approach to improve the prediction of various air pollutants. The model aims to enhance accuracy and efficiency in air quality forecasting by transferring knowledge across different pollutants. The application of transfer learning in air pollution prediction presents a promising method for improving the accuracy of forecasts across multiple pollutants, potentially leading to better environmental management strategies [34]. Another paper uses a one-dimensional multi-scale CNN-LSTM model that considers spatial-temporal features to predict the concentrations of air pollutants in the case of China. It investigates deep learning approaches that integrate spatial and temporal factors in air quality prediction [35]. An alternative study utilizes the LSTM-based neural network model for predicting air pollutant levels. LSTM is recognized for its capability to identify and learn long-term dependencies in time series data. The study evaluates the accuracy and effectiveness of LSTM for air quality prediction and highlights the importance of predictive modeling for air pollution management strategies. It has been a comprehensive study that demonstrates the accuracy and advantages of LSTM models in air pollution prediction [36]. B. Das et al. Their paper discusses using deep learning methods to predict air pollutants in a large metropolitan city. The study applies various deep learning algorithms to analyze air quality dynamics and improve prediction accuracy. It has been an important and valuable study that examines the performance of deep learning methods in air quality assessment in large cities [37]. Another study applies a deep learning-based recurrent neural network (RNN) to predict air pollutants like SO2 and PM10 levels in industrial cities like Sakarya. This research underscores the critical role of predicting air pollution in industrial zones and highlights the capabilities of RNN models for this purpose. The study demonstrates the efficiency and potential of RNN models in forecasting air quality in industrial settings [38]. Furthermore, J. Yang et al. explore spatial and temporal predictions of airborne particle (PM) levels by incorporating data related to traffic and weather conditions. The study aims to improve air quality forecasting by combining different data sources. It highlights the advantages of integrating traffic and weather data to enhance the precision of air pollution predictions [39]. Another study investigates how short- and long-term exposure to air pollutants affects plant phenology and leaf traits. It evaluates the ecological consequences of air pollution on plant health and development. The study provides a review that examines both short- and long-term effects of air pollution on plant biology [40]. In their paper, S. A. Ajayi et al. examine the effects of traffic mobility measures on vehicle emissions under heterogeneous traffic conditions in Lagos. The study analyzes the environmental impacts of traffic regulations and suggests improvement strategies. A significant study has evaluated the emission reduction potential of urban traffic regulations [41]. S. Arslankaya et al. applied machine and deep learning techniques to forecast stock prices. The study evaluates the effectiveness of financial data analysis and forecasting models. It reviews the applications of artificial intelligence techniques for data analysis and forecasting in financial markets [42]. Another study presented a comparative analysis of k-nearest Neighbor (K-NN), Gaussian Naive Bayes (GNB), Support Vector Machines (SVM), Random Forest (RF), and XGBoost models using air quality data from 23 cities in India. The experimental results show that XGBoost performs the best [43]. Baran et al. used an adaptive network-based fuzzy inference system (ANFIS), support vector regression (SVR), classification and regression trees (CART), random forest (RF), K-NN, and extreme learning machine (ELM) methods for the prediction of PM10 and PM2.5 components in Sıhhiye region. The ANFIS model was more successful in predicting PM10 values than other methods [44]. In another study, Dokuz et al. investigated the use of deep learning methods such as Deep Neural Networks (DNN), Recurrent Neural Networks (RNN), and Convolutional Neural Networks (CNN) along with traditional classification algorithms such as LASSO regression, Support Vector Machines (SVM), Random Forest (RF) and K-NN for the prediction of air quality parameters [45]. These studies reveal the potential of various methods and models in air quality prediction and evaluate the effectiveness of different approaches. The existing literature shows that hybrid models, deep learning, and machine learning techniques have a wide application in air quality predictions. These techniques have great potential to increase prediction accuracy. However, each method and model has its limitations and uncertainties, and future studies need to develop more advanced modeling techniques and data integration strategies to address these shortcomings.

In summary, unlike the existing studies in literature, the study includes many innovative and original elements such as comparative analysis of deep learning models, long-term forecasting capacity, use of advanced visualization techniques, local application, use of missing data completion techniques, and detailed explanation of the methods. This highlights the study's important contributions and differences that distinguish it from other studies in the literature.

To provide a more structured comparison of previous studies and highlight this research's innovative contributions, Table 1 summarizes the key studies in the literature. The table focuses on the methods, performance metrics, dataset characteristics, and missing data imputation techniques used. Unlike most previous studies that evaluate a single model or use basic imputation techniques, this study performs a comparative analysis of three deep learning models (RNN, CNN, LSTM) using multiple performance metrics (MAE, RMSE, R^2). The dataset spans 10 years and incorporates advanced missing data imputation via the K-Nearest Neighbor algorithm, offering a more comprehensive approach to air quality prediction.

As seen in Table 1, while previous studies have provided significant insights into air quality prediction, most focus on a single model or limited pollutant types, often without addressing missing data comprehensively. This study distinguishes itself by directly comparing three deep learning models on the same dataset, implementing advanced imputation techniques, and offering long-term forecasts. These methodological advancements are expected to contribute significantly to air quality management and planning.

Table 1. Comparison of Some Air Quality Forecast Methods in Literature

Studies	Model	Pollutant	Performance Metrics	Missing Data Completion	Comparative Analysis
Kurnaz ve Demir [46]	RNN	SO_2	R^2: 0.67	Missing Information	No
Bernardino et al. [47]	Random Forest	SO_2	R^2: -0.0231	No	No
Cerezuela-Escudero [48]	DNN	PM_{10}	R^2: 0.61	Average Filling	No
Kim et al. [49]	LSTM	PM_{10}	R^2: 0.57	No	No
This Study	RNN, LSTM, CNN	PM_{10}, SO_2	R^2: 0.97, 0.97 MAE: 3.52, 1.26 RMSE: 2.47, 0.86	K-Nearest Neighbor	Yes

1.3. Problem Statement and Motivation

Several important points stand out from the articles in the literature in this study.

- Three different deep learning methods for air quality prediction, namely RNN, CNN, and LSTM, were comparatively analyzed on the same dataset. In literature, the performance of a single model is usually evaluated, or the use of a model as a hybrid with another model is prominent. However, in this study, the direct comparison of these three models on the same dataset reveals the advantages and disadvantages of different methods.
- The second difference is that the study used the RNN model to make air quality predictions for the next three years and visualize these predictions. While most studies in the literature make short-term predictions, this approach offers a longer-term forecast. This important contribution can be made when making strategic decisions for air pollution management and planning.
- Another contribution is that it used various advanced visualization methods (such as Box Plot, Violin Plot, and Scatter Plot) to increase the understandability of the prediction results. This approach makes it easier for non-experts to understand and interpret the prediction results. Such techniques are generally limited in literature and usually limited to only basic graphics (such as line graphs).
- Again, many studies in the literature usually make air quality predictions in wide geographical areas or at a general level. However, this study offers a more localized and detailed prediction model by focusing on a specific local area, such as the Başakşehir district of Istanbul. This provides more directly applicable data for local governments to monitor air quality and develop effective intervention strategies.
- In the study, using the K-Nearest Neighbor (K-NN) algorithm to solve the missing data problem is a subject that is missing in many studies in the literature. While the missing data problem is usually approached with simple filling methods (e.g., filling with average), using a more sophisticated method such as K-NN in this study makes a significant difference in increasing the integrity and accuracy of the data.
- In the study, more than one metric (MAE, RMSE, R^2) was used to evaluate the model performance, and the results of these metrics were discussed in detail. While some studies in literature usually focus on a single performance metric, using more than one metric in this study allows a more comprehensive evaluation of the performance of the models.
- Finally, the application steps of each deep learning model used for data preparation, model architecture, and the parameters used are explained in great detail. Such details are usually briefly mentioned in the literature, and it may be difficult for the reader to understand the subject fully. However, this approach reveals how the model works and why it was chosen in this way.

The remainder of the paper is structured as follows: Section 2 presents general information about data visualization methods, missing data completion techniques, and deep learning models used in air quality data collection. Section 3 details the proposed methodology, explaining the dataset, preprocessing steps, the K-Nearest Neighbor (K-NN) algorithm used for missing data completion, and the prediction processes performed with deep learning models, CNN, RNN, and LSTM. Section 4 presents experimental results and discussions. The success rates of CNN, RNN, and LSTM models in air pollution predictions were evaluated using metrics like MAE, RMSE, and R^2, and the most successful model was determined. Graphical results are presented with methods such as violin plots, box plots, and scatter plots. In addition, the next 3-year prediction results are discussed. Section 5 represents a general evaluation of the study and the conclusions. It was concluded that the RNN model provided the most successful results in air quality predictions by learning both short-term and long-term dependencies. At the end of the study, forward-looking suggestions were made, and it was stated that model performance could be improved by adding meteorological data.

2. Background

This part of the study offers an overview of data visualization techniques, methods for handling missing data, and deep learning approaches.

2.1. Data Collection

Accurate and reliable data collection forms the basis of the research and analysis processes. The data used in this study are air quality indices (AQI) data collected from various environmental sensors. At certain intervals, the sensors measure various air quality parameters (PM2.5, PM10, SO2, NO2, CO, O3). This collected data can be recorded in databases and made available for analysis. However, incomplete data may be encountered during the collection for reasons such as sensor failures, connection problems, or environmental factors. This incomplete data must be completed and processed with appropriate methods to perform accurate analysis.

2.2. Filling in Missing Data Method

One of the most frequent challenges when working with data is handling missing data. Missing values can negatively impact the accuracy and validity of analysis and visualization. Various algorithms and methods can be applied to fill in missing data to address this. K-NN is a method commonly used for classification and regression in supervised learning. In the context of missing data imputation, K-NN replaces the missing value with the average of the 'k' nearest neighbors' values.

In this study, the K-Nearest Neighbor approach was chosen to address missing data issues due to its ability to maintain the local structure of the dataset and preserve critical interdependencies among features. Air quality datasets, which often include PM2.5, PM10, SO2, and NOx, inherently exhibit complex multivariate relationships and spatial-temporal dependencies. Unlike simpler imputation techniques such as mean or median substitution, which assume a uniform distribution of missing values and fail to account for these dependencies, K-NN uses similarity-based criteria to identify and use the most relevant neighboring data points for imputation. This ensures that the imputed values align with the dataset's statistical properties and reflect the localized patterns and trends inherent in the data. Additionally, preliminary tests conducted during the data preprocessing phase demonstrated that K-NN outperformed simpler methods in terms of both error metrics and downstream predictive model performance. Specifically, K-NN-imputed data resulted in lower mean absolute error (MAE) and root mean square error (RMSE) values than mean substitution. In contrast, models trained on K-NN-imputed datasets exhibited better accuracy and generalization capabilities. This empirical evidence supports the notion that K-NN provides a more robust and contextually appropriate solution for missing data imputation, particularly in applications involving air quality estimation, where data quality directly influences the reliability of forecasts [43].

2.3. Data Fusion and Gaussian Filter

Data fusion means using air quality data from sensors to make more comprehensive and reliable air quality estimates. Data fusion increases data accuracy and predictive power by integrating data from different values. This method contributes to more reliable analysis results. Gaussian filter is a data filtering method used to reduce the noise of environmental data and make the signal smoother. Gaussian filter averages the data by weighing it according to a certain distribution, thus obtaining a more stable and regular data set. This study reduced the noise in air quality data, and more accurate estimates were obtained using the Gaussian filtering method.

2.4. Data Visualization Methods

The process of simplifying large and intricate data sets to make them easier to understand and meaningful by visually representing them through graphs is known as data visualization. Data visualization converts raw data from challenging numerical sets into visual formats where trends, relationships, and patterns can be easily identified. This study used three data visualization techniques to analyze and interpret the predicted data better.

One of these, the Point Scatter Plot, is a technique utilized to analyze the relationship between two different numerical data. This graph visualizes possible correlations or distribution trends between the data, allowing researchers to understand the data more deeply. This method visualizes the relationship between two variables and can show many values in a data set through different colors or sizes. Point scatter plots determine possible correlations or patterns between data and make complex data more understandable [50].

Box Plot is a method used to visualize the distribution, central tendency, and possible data outliers. It makes it easier to understand the general distribution of the data by showing the minimum, maximum, median, first, and third quartiles in a data set. The box plot is especially useful for visualizing comparisons between different data groups and effectively determining possible extreme values in the data set.

The violin plot illustrates the data distribution like the box plot, but it also presents the density distribution of the data. A violin plot provides more comprehensive details regarding the data's form and distribution by visualizing the data's probability density function. This approach enables us to analyze the symmetry and distribution of the data sets in more detail [51].

3. Material and Methods

This study gathered a 10-year dataset from 2014 to 2024 from the Continuous Monitoring Center, affiliated with Turkey's Ministry of Environment, Urbanization, and Climate Change. Afterward, missing data were identified and imputed using K-Nearest Neighbor (K-NN) algorithms. Finally, three deep learning methods, CNN, RNN, and LSTM, were applied to the data completed with k-nn. The results were compared, and the most accurate values were obtained using RNN. After this process, the RNN method was selected, and the next 3 years were estimated. Then, a point distribution graph, box plot, and violin graph were used in the data, respectively. The diagram showing the application steps is shown in Figure 3.

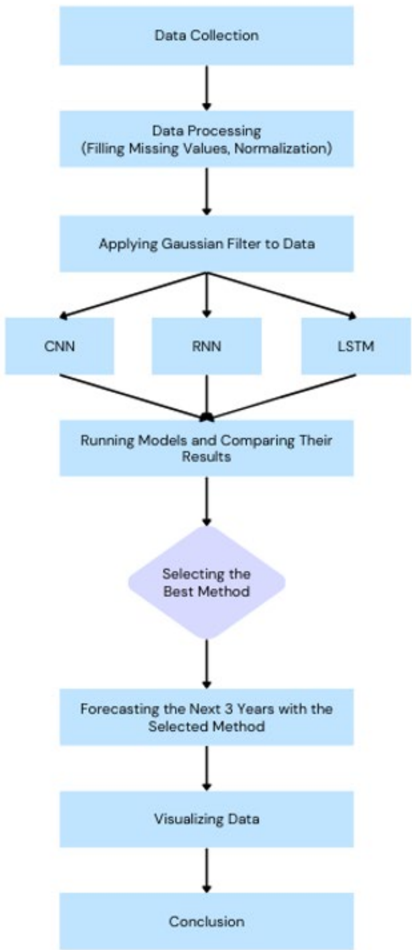

Figure 3. Diagram Outlining the Study

3.1. Dataset

This study obtained data from the Continuous Monitoring Center, made publicly available by the Ministry of Environment, Urban Planning, and Climate Change of the Republic of Turkey [52]. This dataset, covering the years 2014-2024, served as the foundation for the research. The dataset includes relevant air pollutant values and date information for each day and is presented ready for analysis. The first 10 rows of the original data collected are given as an example in Table 2.

The dataset comprises 3655 rows of records in total, and the missing data rates are remarkable:

- Missing PM10 Data: 246 rows
- Missing SO2 Data: 364 rows

- Missing CO Data: 376 rows
- Missing NO2 Data: 224 rows
- Missing NOX Data: 542 rows
- Missing NO Data: 335 rows
- Missing O3 Data: 255 rows

The main reason for choosing this comprehensive 10-year dataset is to train models more effectively and powerfully with deep learning methods. A large dataset enhances the model's sensitivity and significantly improves the accuracy of predictions in environmental analyses. Large data sets improve the model's performance and provide an important advantage for increasing the accuracy of the analysis. For the experimental analysis, 64% of the data was allocated for training, 16% for validation, and 20% was reserved for testing. This division ensures that the model is well-trained in sufficient data, properly validated during training, and adequately evaluated on unseen data to assess its performance effectively.

Table 2. The First 10 Rows of the Original Dataset

Date	**PM10(μg/m3)**	**SO2(μg/m3)**	**NO2(μg/m3)**	**CO(μg/m3)**	**NO(μg/m3)**	**NOX(μg/m3)**	**O3(μg/m3)**
2014-03-04 00:00:56	70,57	8,54	49,56	460,87	nan	nan	35,23
2014-03-05 00:00:56	135,89	23,23	65,74	756,52	nan	nan	19,28
2014-03-06 00:00:56	53,92	4,50	37,21	217,39	nan	nan	47,35
2014-03-07 00:00:56	52,29	nan	26,60	273,91	nan	nan	62,34
2014-03-08 00:00:56	21,05	nan	15,78	352,17	nan	nan	56,39
2014-03-09 00:00:56	23,71	nan	8,89	369,57	nan	nan	63,58
2014-03-10 00:00:56	21,43	nan	15,80	343,48	nan	nan	59,63
2014-03-11 00:00:56	26,06	nan	14,01	527,27	nan	nan	67,79
2014-03-12 00:00:56	nan	nan	nan	nan	nan	nan	nan
2014-03-13 00:00:56	nan	nan	nan	nan	nan	nan	nan

3.2. Preprocessing

The K-Nearest Neighbor (K-NN) algorithm offers a highly effective approach to complete missing data. This algorithm estimates a value with missing data by looking at its similarities with other values in the dataset. The K-NN algorithm determines the 'k' nearest neighbors to complete a missing value in the dataset and estimates the missing values using the average of these neighbors. The K-NN algorithm not only completes the data but also increases the accuracy and integrity of the dataset, allowing it to work effectively in large datasets.

K-NN was chosen for this study because, in our previous work, we used the mean to address missing data [43]. However, we found that using the mean was insufficient for accurate prediction after completing the missing data. As a result, we opted for K-NN, which demonstrated superior performance in handling missing values and improving predictive accuracy. Using K-NN offers an effective solution, especially in estimating missing values in the dataset with variable values, such as the 10-year air quality in our study. Table 3 below shows the first 10 rows of the dataset filled with K-NN.

Table 3. Dataset Filled with K-NN

Date	**PM10(μg/m3)**	**SO2(μg/m3)**	**NO2(μg/m3)**	**CO(μg/m3)**	**NO(μg/m3)**	**NOX(μg/m3)**	**O3(μg/m3)**
2014-03-04 00:00:56	70.57	8.54	49.56	460.87	29.41000	100.948	35.23
2014-03-05 00:00:56	135.89	23.23	65.74	756.52	67.178	198.724	19.28
2014-03-06 00:00:56	53.92	4.5	37.21	217.39	21.892	66.328	47.35
2014-03-07 00:00:56	52.29	28.74	26.6	273.91	10.978	40.892	62.34
2014-03-08 00:00:56	21.05	21.534	15.78	352.17	6.196	24.17	56.39
2014-03-09 00:00:56	23.71	3.304	8.89	369.57	4.672	20.01	63.58
2014-03-10 00:00:56	21.43	12.302	15.8	343.48	6.51	25.16	59.63
2014-03-11 00:00:56	26.06	2.686	14.01	527.27	3.196	19.578	67.79
2014-03-12 00:00:56	48.78043	7.12803	28.35337	517.94021	14.91725	52.35370	56.82346
2014-03-13 00:00:56	48.78043	7.12803	28.35337	517.94021	14.91725	52.35370	56.82346

3.3. Methods Used and Analysis

This paper used three deep learning methods, namely CNN, RNN, and LSTM, to predict air pollutant data. During the application of these methods, the data was first organized with a Gaussian filter, and predictions were made for each method. The outcomes were assessed based on R^2 scores, Root Mean Square Error (RMSE), and Mean Absolute Error (MAE). Finally, these studies were supported with graphics.

MAE represents the average of the absolute differences between predicted and actual values. This metric measures the magnitude of prediction errors and assigns equal weight to all errors. The Formula 1 for MAE is as follows:

$$MAE = \left(\frac{1}{n}\right) \Sigma \left| y_i - \bar{y}_i \right| \qquad \text{Formula 1}$$

Where:

- y_i: Actual value,
- $\bar{y}$: Predicted value,
- n: Total number of data points.

MAE is simple to interpret and directly measures the average error size. However, it does not emphasize larger deviations, making it less sensitive to extreme prediction errors than other metrics.

RMSE calculates the square root of the average squared differences between predicted and actual values. Squaring the errors assigns greater importance to larger deviations, offering a better understanding of error distribution. In Formula 2 for RMSE is:

$$RMSE = \sqrt{\left(\frac{1}{n}\right) \Sigma \left(y_{\mathrm{i}} - \bar{y}_{\mathrm{i}}\right)^{2}} \qquad \text{Formula 2}$$

Where:

- y_{i}: Actual value,
- $\bar{y}_{\mathrm{i}}$: Predicted value,
- n: total number of data points.

RMSE is particularly useful when large prediction errors are critical to the analysis. It provides a more comprehensive view of the model's performance by emphasizing significant deviations. However, it is slightly more complex to interpret compared to MAE.

In Formula 3, R^2 measures the proportion of variance in the dependent variable that the model explains. It provides an overall evaluation of how well the model captures the data's variability. The formula for R^2 is:

$$R^{2} = 1 - \left(\Sigma \left(y_{\mathrm{i}} - \bar{y}_{\mathrm{i}}\right)^{2}\right) / \left(\Sigma \left(y_{\mathrm{i}} - \bar{\bar{y}}\right)^{2}\right) \qquad \text{Formula 3}$$

Where:

- y_{i} : Actual value,
- $\bar{y}_{\mathrm{i}}$: Predicted value,
- $\bar{\bar{y}}$: Mean of actual values,

R^2 values typically range from 0 to 1, where 1 indicates perfect prediction, and 0 indicates no explanatory power. Negative R^2 values suggest the model performs worse than a simple mean-based prediction. This metric is particularly useful for assessing the model's overall fit to the data. These three metrics collectively provide a holistic evaluation of model performance. MAE focuses on the average error magnitude, offering a straightforward interpretation. RMSE emphasizes larger errors, making it more informative for analyzing error distribution and model reliability. R^2 evaluates the model's ability to explain variance in the target variable, providing a summary measure of its overall effectiveness.

This study employs MAE, RMSE, and R^2 to compare the predictive performance of CNN, RNN, and LSTM models. MAE and RMSE quantify prediction errors, while R^2 assesses the models' explanatory power. This combination of metrics ensures a comprehensive analysis of the models' accuracy and generalization capacity for forecasting PM10 and SO2 levels.

For air pollutant modeling, performance metrics like R^2 , MAE, and RMSE are crucial indicators of model accuracy. An R^2 value between 0.9 and 1 is considered ideal, indicating the model explains most of the variability in the data, while values between 0.8 and 0.9 are still acceptable. For MAE, values below 5 µg/m³ are ideal, with 5–10 µg/m³ being acceptable, as lower errors mean predictions are closer to actual values. Similarly, RMSE should ideally be below 7 µg/m³, with values between 7 and 12 µg/m³ still considered good. These thresholds ensure reliable predictions for pollutants like PM10 and SO_2, aligning with air quality standards.

A Gaussian filter is a data filtering application that reduces the noise of environmental data and makes the signal smoother. This application takes the average of the data by weighting it according to a certain distribution, and thus, a more stable data set is obtained.

Convolutional Neural Networks (CNNs), while primarily utilized for image processing tasks, have proven to be highly effective for time series data analysis [53]. Their ability to extract localized features and learn short-term dependencies makes them suitable for sequential data. In this study, the CNN architecture, illustrated in Figure 4, has been optimized for forecasting PM10 and SO2 levels.

The model begins with a data input layer, where raw time series data is fed into the network. The first layer is a one-dimensional convolutional (Conv1D) with 64 filters and a kernel size of 2. This layer slides a filter over the data, extracting local features and patterns essential for understanding the structure of the time series. Following this, a MaxPooling1D layer with a pooling size of 2 reduces the dimensionality of the data. The model retains critical information while discarding unnecessary details, creating a more compact representation.

A dropout layer with a rate of 0.2 is included to prevent overfitting. This layer randomly disables some neurons during training, ensuring the model generalizes well to new data. The data then passes through a flattening stage, where the multi-dimensional representation is converted into a one-dimensional vector, preparing it for the dense layers.

The dense layers form the final stage of the model. A dense layer with 50 neurons learns high-level representations of the data. Subsequently, an output layer with two neurons generates predictions for the target variables, PM10 and SO2. The model is optimized using the Adam optimizer, with a learning rate of 0.001, and employs the mean squared error loss function.

This CNN architecture is especially adept at processing time series data, leveraging its capability to learn and predict patterns effectively. This study demonstrates that it provides a robust framework for tasks like forecasting air quality parameters.

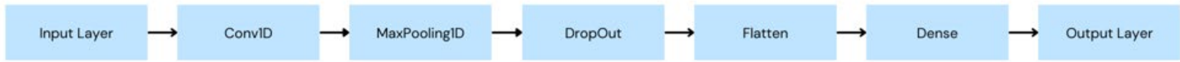

Figure 4. CNN Model Architecture

Recurrent Neural Networks (RNNs) are specialized artificial neural networks designed to model sequential data, such as time series [54]. Their primary advantage lies in their ability to retain information from previous time steps in a hidden state, enabling them to capture temporal dependencies effectively. This makes RNNs highly suitable for tasks like forecasting and sequential data analysis.

In this study, the RNN architecture, depicted in Figure 5, is tailored for predicting PM10 and SO2 levels. The model starts with an input layer where the sequential data is received and prepared for processing. The first layer is a SimpleRNN layer with 50 neurons. This configuration ensures that the outputs of the first RNN layer are passed to the next layer, allowing the model to learn dependencies across multiple time steps. Following the first RNN layer, a dropout layer with a rate of 0.2 is applied. This layer randomly deactivates neurons during training, reducing the risk of overfitting and improving the model's generalization ability.

The architecture continues with a second SimpleRNN layer containing 50 neurons, which deepens the model's ability to capture sequential patterns. Another dropout layer with a rate of 0.2 is added after this RNN layer to enhance the model's robustness and prevent over-learning.

Finally, the data is processed by an output layer of two neurons, generating predictions for PM10 and SO2 concentrations. The model is trained using the Adam optimizer with a learning rate of 0.001, and the mean squared error loss function is employed to minimize prediction errors.

This RNN architecture leverages its capacity to store and utilize past information, making it a powerful tool for time series prediction. The model achieves accuracy and generalization in its predictions by combining sequential learning with dropout regularization.

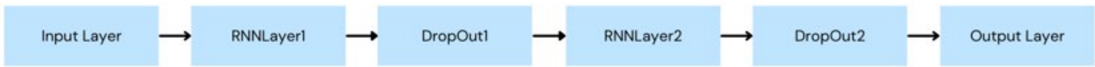

Figure 5. RNN Model Architecture

Long Short-Term Memory (LSTM) networks are a specialized variant of Recurrent Neural Networks (RNNs) that excel at capturing both short-term and long-term dependencies in sequential data [55]. Their unique cell state and gating mechanisms allow them to effectively learn patterns across extended time steps, making them particularly suitable for time series forecasting tasks.

In this study, the LSTM architecture, illustrated in Figure 6, is designed to predict PM10 and SO2 concentrations. The model begins with an input layer where the sequential data is introduced. The first processing stage involves an LSTM layer with 50 neurons. This configuration enables the first LSTM layer to output sequential data passed to the next layer for further processing. Following this, a dropout layer with a rate of 0.2 is applied to randomly deactivate neurons, reducing the risk of overfitting and improving generalization.

The second stage consists of another LSTM layer with 50 neurons, which deepens the model's ability to learn complex dependencies in the data. Similar to the first stage, a dropout layer with a rate of 0.2 is applied after the second LSTM layer to enhance robustness further. Finally, the output layer, containing two neurons, generates predictions for the target variables, PM10 and SO2.

The model is optimized using the Adam optimizer with a learning rate of 0.001, and the mean squared error loss function is used to minimize prediction errors. During training, EarlyStopping is implemented to monitor the validation loss, with a patience of 100 epochs. This ensures that training halts if performance does not improve, preventing unnecessary computations and overfitting. Additionally, ModelCheckpoint is utilized to save the best-performing model for future use.

This LSTM architecture effectively captures temporal dependencies in the data, leveraging its advanced structure to deliver accurate and generalizable predictions for time series forecasting tasks.

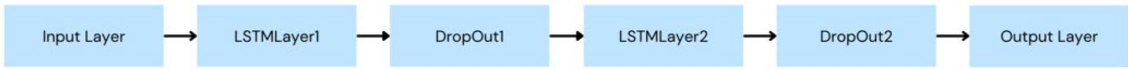

Figure 6. LSTM Model Architecture

The performance of the models is evaluated using three key metrics: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and the R^2 score. These metrics collectively provide a comprehensive assessment of the models' predictive capabilities. Mean Absolute Error (MAE) represents the average of the absolute differences between the predicted and actual values. It serves as a straightforward measure of prediction accuracy by quantifying the magnitude of errors, without considering their direction. Root Mean Square Error (RMSE) is calculated as the square root of the average squared differences between predicted and actual values. This metric gives greater weight to larger errors, making it particularly sensitive to significant deviations in predictions. R^2 Score evaluates the proportion of the variance in the dependent variable that is explained by the model. It provides an overall indication of how well the model captures the variability in the target data, with a higher R^2 score indicating better predictive performance. By comparing these metrics across the CNN, RNN, and LSTM architectures, the study identifies the most effective model for forecasting PM10 and SO2 time series data. Each metric contributes unique insights into the models' performance, enabling a thorough evaluation of their strengths and limitations.

Table 4. Comparison of Hyperparameters for CNN, RNN, and LSTM Models

Hyperparameter	**CNN**	**RNN**	**LSTM**
Model Nodes	64 CNN Filters	64 RNN Nodes	2 LSTM Nodes
Epoch	20, 30, 40	20, 30, 40	20, 30, 40
Batch Size	16	16	16
Interpolate Method	N/A	N/A	Linear
Train Data	64% Dataset	64% Dataset	64% Dataset
Validation Data	16% Dataset	16% Dataset	16% Dataset
Test Data	20% Dataset	20% Dataset	20% Dataset
Optimizer	SGD	ADAM	ADAM
Learning Rate	0.0001	0.001	0.001
Dense Layer	64	N/A	N/A

The hyperparameter Table 4 of the study is given above. This table shows the comparatively determined parameters of CNN, RNN, and LSTM models. While the CNN model exhibits a structure suitable for image processing, it worked with 64 filters and the SGD optimization method. RNN and LSTM models are designed for stronger performance on time series data and trained with ADAM optimizer. LSTM successfully captured long-term dependencies using the "linear" interpolation method, while RNN was more effective in learning short-term and long-term relationships. All models were run on the same data rates (64% training, 16% validation, 20% test) and similar epoch values. These hyperparameters were carefully selected to evaluate how the models respond to different data structures.

4. Experimental Results and Discussion

In this study, analyses were conducted on PM10 and SO2 pollutants using CNN, LSTM and RNN as air quality prediction models. In the study, the performance of these models was evaluated with metrics such as MAE (Mean Absolute Error), RMSE (Root Mean Square Error) and R^2. According to the results, the RNN model showed the best performance with the lowest error rates. The capacity of RNN to learn short- and long-term dependencies enabled the model to be more successful in predictions. While LSTM was successful in capturing long-term dependencies, it did not perform as well as RNN. The CNN model had higher error rates and had difficulty in capturing high variances, especially in time series data. According to the experimental results, the RNN model achieved the highest accuracy in PM10 and SO2 predictions, while CNN exhibited lower performance with higher error values. The differences between the modeled predictions and real values were visualized with violin plot and scatter plot graphics, and higher air pollution was predicted in certain years, especially in PM10 predictions.

Among the models used in this study, accuracy rates of 0.88 for PM10 and 0.93 for SO2 were achieved with CNN; 0.94 for PM10 and 0.95 for SO2 were achieved with LSTM. The most successful model, RNN, provided an accuracy rate of 0.97 for PM10 and SO2, and forecasts are made for the next 3 years using this model. The values and visualizations resulting from the study are given in detail below. The CNN model results shown in Table 5 provide reasonable accuracy in PM10 and SO2 forecasts. However, the MAE and RMSE values are higher than the other models, especially for PM10, indicating that the errors are larger. The R^2 scores are at the levels of 0.88 and 0.93, indicating that the model's forecast performance is generally good, but it has higher error rates.

Table 5. CNN Model Performance Metrics

Pollutants	MAE	RMSE	R^2
PM10	4.68	6.44	0.88
SO2	1.34	2.07	0.93

The RNN model in Table 6 has the lowest MAE and RMSE values in both PM10 and SO2 predictions, showing the best performance. This shows that the RNN model can successfully learn both short-term and long-term dependencies and make more accurate predictions. The R^2 scores are also at 0.97, which shows that the model explains the variance very well and the predictions are very close to the actual values.

Table 6. RNN Model Performance Metrics

Pollutants	MAE	RMSE	R^2
PM10	2.47	3.52	0.97
SO2	0.86	1.26	0.97

The LSTM model in Table 7 performs slightly lower than the RNN but still shows a good performance. The MAE and RMSE values are higher than the RNN but lower than the CNN. The R^2 scores are 0.94 and 0.95, demonstrating that the model explains a significant portion of the dataset's variability. While the LSTM performs well in identifying and capturing long-term dependencies in time series data, it is less effective than the RNN in this particular scenario.

Table 7. LSTM Model Performance Metrics

Pollutants	MAE	RMSE	R^2
PM10	3.67	4.66	0.94
SO2	1.07	1.85	0.95

As a result, the LSTM model gives a result close to RNN in terms of its performance on time series data, but it does not perform as well as RNN. Nevertheless, it made better predictions compared to CNN. The fact that the CNN model has higher error values indicates that its capacity to capture serial dependencies in time series data is lower than other models. The RNN model has the lowest error values (MAE and RMSE) and the highest R^2 scores in both PM10 and SO2 predictions and stands out as the most successful model for this data set and problem. The study provides better results than other studies conducted before [56-58]. In Figure 7, which represents the performance of the CNN model for predicting PM10 and SO_2 concentrations, there is a noticeable difference between the actual values (blue and orange lines) and the predicted values (cyan and red lines). These differences provide insights into the model's strengths and weaknesses.

For PM10 predictions, the model generally captures the overall trends of the real data, especially during periods of lower concentration. However, significant deviations are observed during peaks, particularly in high pollution events. These deviations suggest that the CNN model struggles to accurately predict extreme values, which could be due to the limited ability of CNNs to model abrupt changes or anomalies in time series data. Despite this, the model successfully tracks seasonal and periodic fluctuations.

For SO_2 predictions, the CNN model performs better at capturing the general trends compared to PM10. The predicted values (red line) closely align with the actual values (orange line) during periods of stability. However, similar to PM10, the model shows weaknesses in predicting sudden spikes or drops in SO_2 levels. This may indicate that while CNNs are effective at identifying overall patterns, they may require additional features or architectural modifications to handle abrupt changes more effectively.

Additionally, the graph highlights that for both pollutants, the predicted values show a slightly smoother pattern compared to the real values. This smoothing effect is common in CNN models, as they prioritize extracting dominant trends rather than capturing noise or highly localized variations. While this improves the generalization of the model, it can reduce its ability to capture sharp fluctuations accurately. In summary, the CNN model demonstrates the ability to follow the overall patterns and seasonal variations of both PM10 and SO_2 levels.

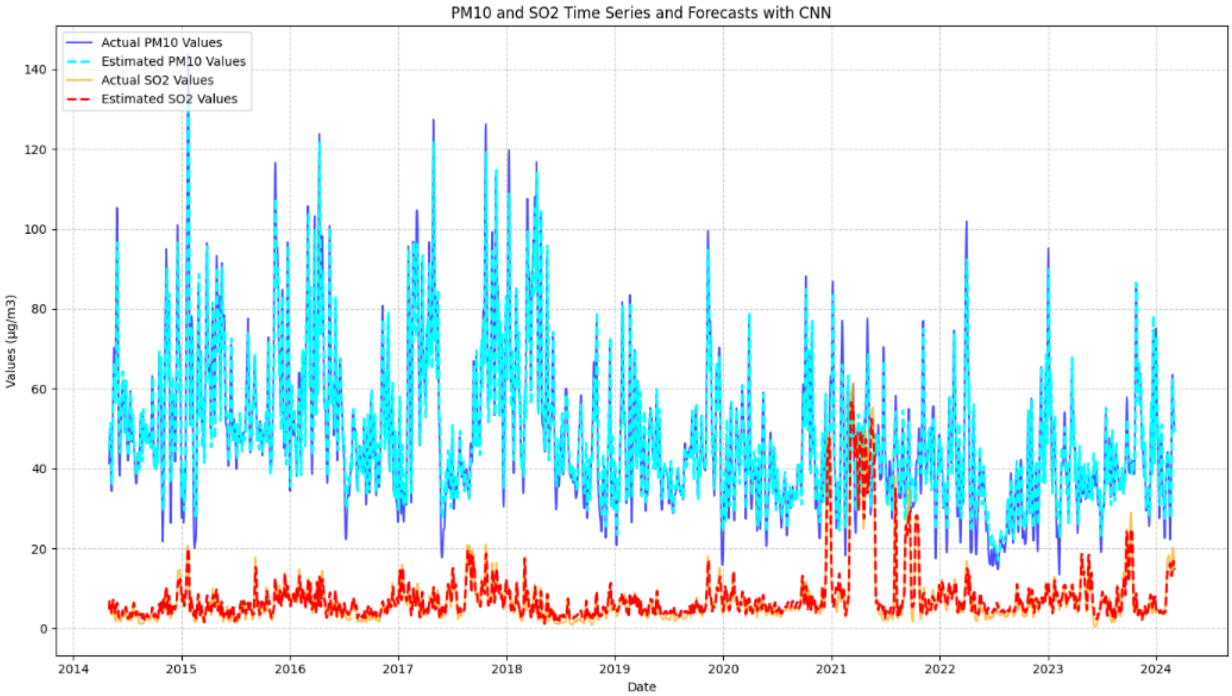

Figure 7. Prediction with CNN Model

In the graph in Figure 8, the performance of the RNN model is demonstrated, showing an impressive alignment between the predicted values (cyan and red lines) and the actual values (blue and orange lines). The RNN model excels at capturing the variations and fluctuations in the time series data for both PM10 and SO_2, with minimal differences between predictions and real values compared to other models.

For PM10 predictions, the predicted values (cyan line) closely follow the actual data (blue line), successfully capturing both gradual and sudden changes. This is especially evident in high pollution events, where the RNN model handles the peaks better than other models, resulting in smaller deviations during extreme conditions. The alignment in seasonal and periodic patterns further demonstrates the model's strength in learning temporal dependencies.

For SO_2 predictions, the predicted values (red line) show an even stronger agreement with the actual values (orange line). The model accurately tracks both the general trends and the sharp fluctuations in SO_2 concentrations. Sudden increases and decreases are effectively modeled, showcasing the RNN's ability to handle dynamic changes in time series data.

The graph also reveals the RNN model's ability to generalize across different periods. Unlike other models, the prediction lines do not exhibit significant smoothing, indicating that the model effectively preserves the detailed patterns and variability present in the data. This allows the RNN to achieve the lowest MAE and RMSE values among the compared models, making it the most accurate in predicting both PM10 and SO_2 levels. Overall, the RNN model's capability to capture both short-term and long-term dependencies in the time series data is evident. This makes it a reliable choice for air quality prediction tasks, especially when accurate forecasting of abrupt changes and complex patterns is required.

In Figure 9, the LSTM model's performance is depicted, showing strong agreement between the predicted values (cyan and red lines) and the actual values (blue and orange lines) for both PM10 and SO_2 levels. This indicates that the LSTM model is effective in capturing temporal dependencies and producing reliable predictions.

For PM10 predictions, the model demonstrates a good fit, closely following the actual data. The predicted values (cyan line) align well with the actual values (blue line) during both periods of stability and fluctuations. While the model captures seasonal patterns effectively, it occasionally underestimates extreme peaks. However, these deviations are relatively small, and the model retains a good balance between generalization and precision, making it comparable to the RNN in performance.

For SO_2 predictions, the LSTM model also shows a strong alignment with the actual values (orange line). The predicted values (red line) follow the overall trends effectively, accurately modeling gradual changes and periodic behaviors. However, the model struggles slightly with sudden spikes or sharp drops in SO_2 levels, leading to minor deviations. Despite these occasional mismatches, the model maintains consistency in capturing broader trends. Overall, the LSTM model handles both pollutants effectively, with its performance close to the RNN model. While the MAE and RMSE values are slightly higher, LSTM excels in capturing long-term dependencies within the time series data. Its ability to balance accuracy and generalization makes it a strong candidate for modeling and forecasting air pollutant levels, particularly in scenarios requiring the identification of broader trends over extended periods.

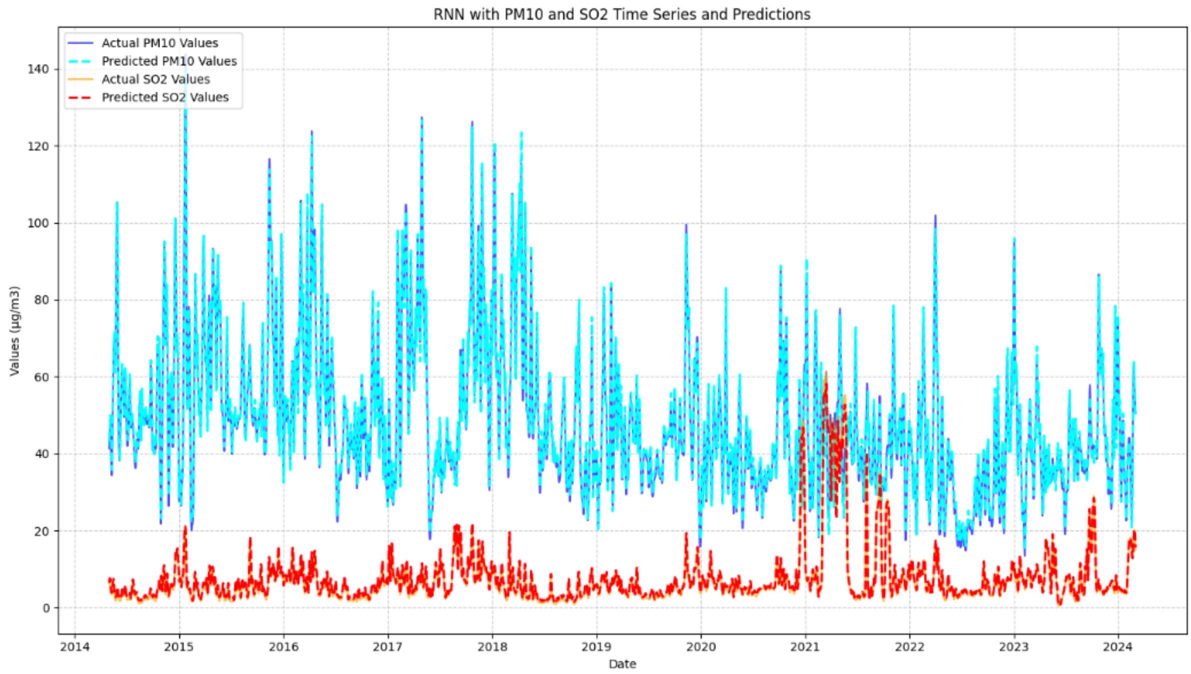

Figure 8. Prediction with RNN Model

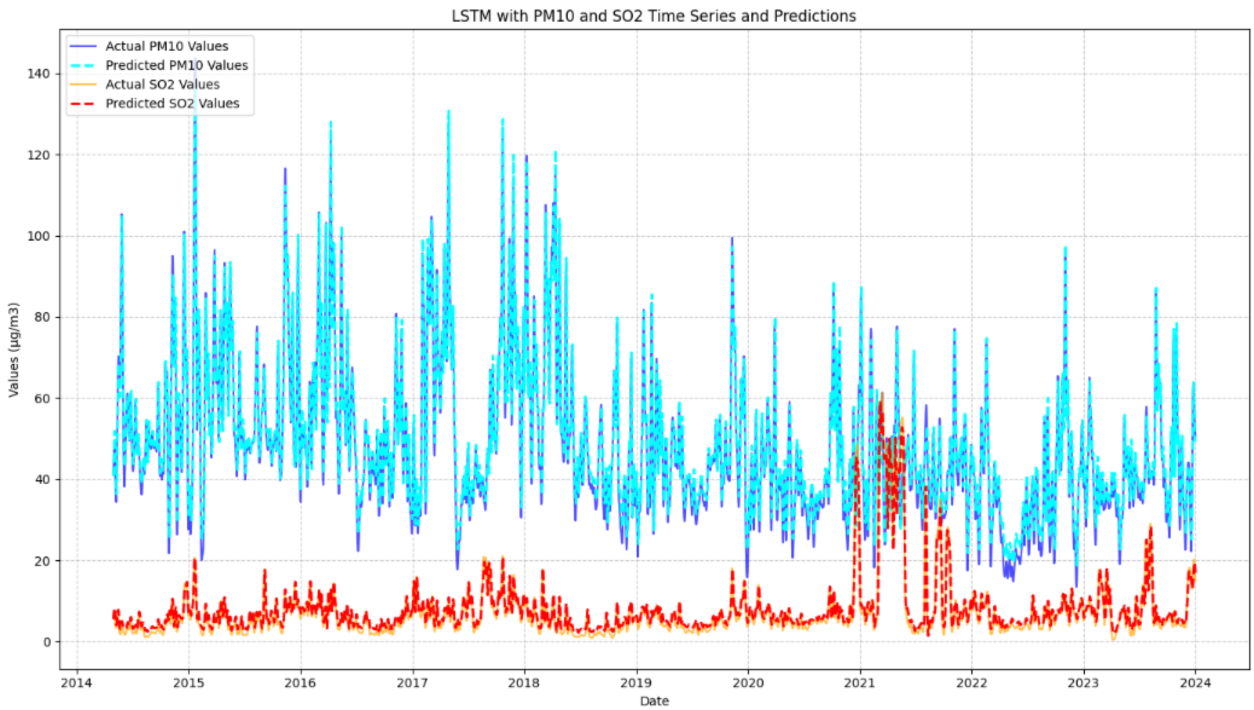

Figure 9. Prediction with CNN Model

As a result, the model that showed the best performance was RNN. The discrepancies in the differences between the actual and predicted values are minimal and the model effectively learns both short and long-term relationships or dependencies within the data. In the second place, LSTM has a good performance and is close to RNN. It has been successful in capturing long-term dependencies in the time series. CNN has a lower performance compared to other models. It has difficulty in capturing the high variance of PM10 values in particular. CNN, which is more suitable for image processing, is not as effective as other models in time series data. Since our study is the most successful RNN model among these three methods, we continue with this method when making future predictions. The output of the performance metrics for the next 3 years is shown in Table 8 below. According to this table, it indicates that the model's predictions are not perfect compared to the actual data but show a moderate performance. It shows that the model can make larger errors in some of its predictions, but it is generally at a reasonable level of accuracy.

Table 8. Performance Metrics in Future Prediction with RNN Model

Pollutants	**MAE**	**RMSE**	**R²**
PM10	11.60	17.15	0.63
SO2	2.39	3.99	0.82

In the graphical output of the model in Figure 8, the future predictions for PM10 are flatter and exhibit lower fluctuations. This indicates that the model may have limited capacity to predict future changes. For SO2, the predictions are relatively lower variance and follow a flatter course. Flatter and lower fluctuations are observed for both PM10 and SO2, indicating that the model may have limited capacity to manage uncertainties and may have difficulty predicting more complex events.

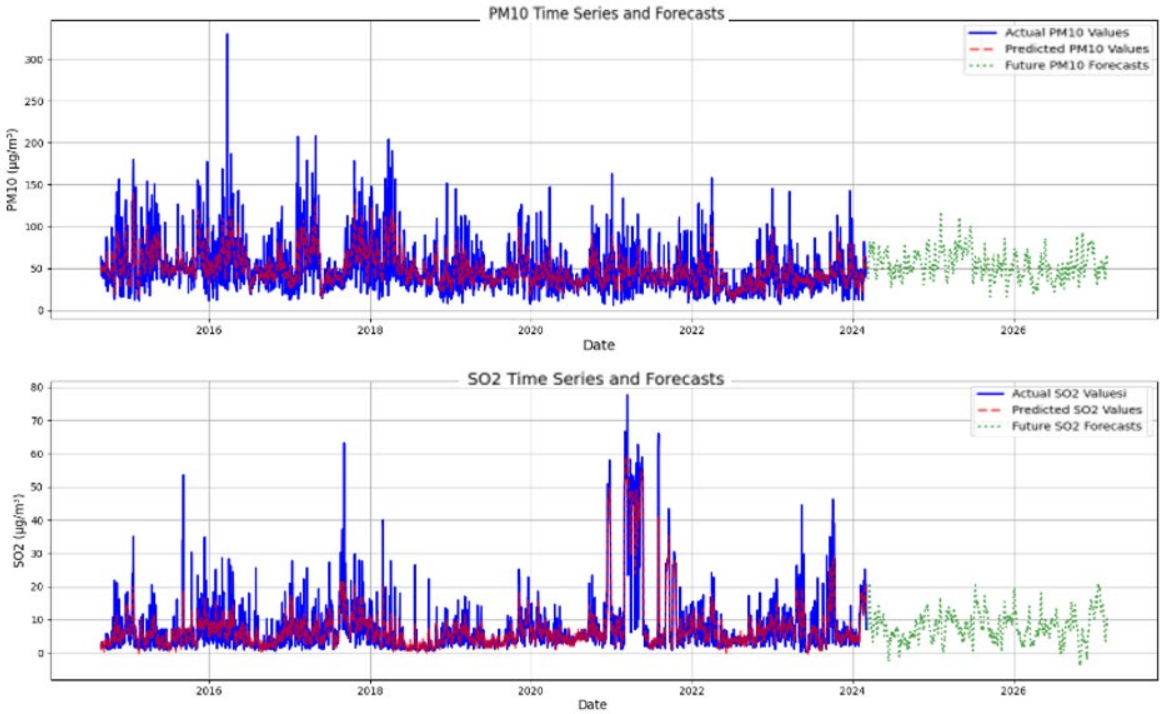

Figure 8. Next 3 Years Prediction Graph with RNN

As a result, the model performs moderately for PM10, while it is more successful in SO2 predictions. The fact that future predictions are relatively flat and have low variance reveals some limitations of the model's predictive capacity. These results can be used for future air quality management and planning, but they indicate that the predictions should be evaluated carefully. Further developments and different model structures can increase the prediction accuracy.

In the final stage of the study, efforts were made to make this large dataset more manageable and complex data easier and more understandable. For this purpose, data visualization was made in the study. To make comparison easier, available data and data obtained with future predictions were used.

Figure 9 summarizes the changes in PM10 and SO2 levels over the years and their future predictions. While PM10 levels showed a wider distribution in 2021 and 2024, indicating that air pollution increased in these years, they concentrated in narrower and lower concentrations in 2022 and 2023, indicating relatively better air quality. PM10 projections for 2024-2027 show a wider distribution than in previous years, indicating that the model has uncertainty in predicting future PM10 levels. SO2 levels, on the other hand, generally show consistent distribution, but in some years, such as 2021, they show a wider distribution, indicating periods of poorer air quality. Low concentrations of SO2 in 2020, 2022, and 2024 indicate cleaner air conditions in these years. Future SO2 projections, although generally concentrated at low levels, show a wider distribution in 2025 and 2027, indicating that the model has more uncertainty for these years.

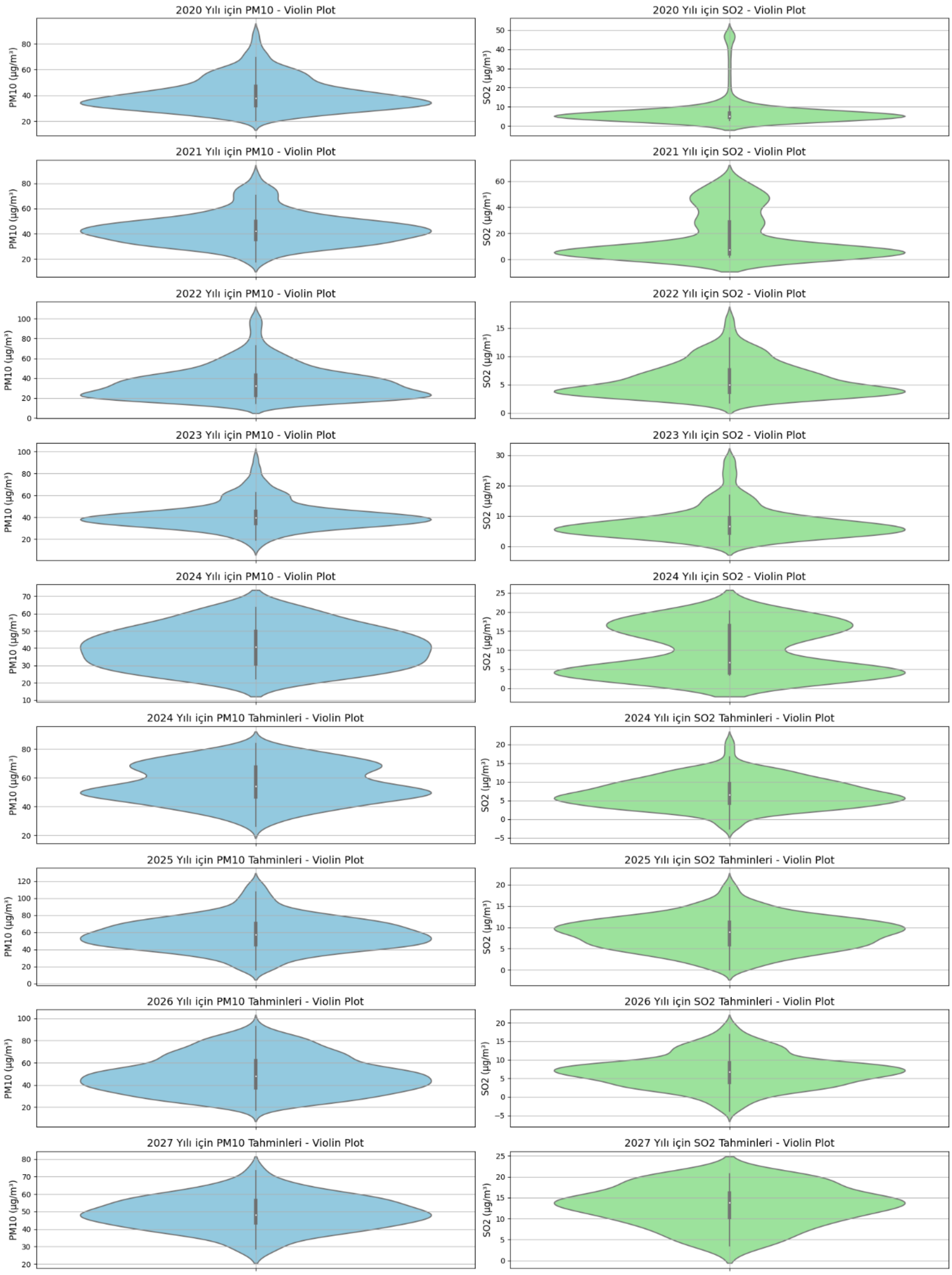

Figure 9. Visualization with Violin Plot

When the actual data between 2020-2024 in Figure 10 is analyzed, it becomes evident that PM10 levels contain more extreme values in some years (especially 2021 and 2023) and air pollution events are more frequent and intense in these years. It is understood that the median values for PM10 are generally concentrated between 20-40 µg/m³ and the variance between years is generally similar. SO2 levels generally remain in a low range (2-10 µg/m³), but more variance and some extreme values are noted in 2021 and 2022. This shows that air quality varies especially in these years and that unexpectedly high SO2 concentrations are experienced in some periods. When looking at future projections, the distribution of PM10 and SO2 levels predicted by the model between 2024-2027 seems generally consistent with previous years but contains some uncertainties and variances. For PM10 estimates, the wider bins (IQR) indicate that the model predicts more variance in future PM10 levels and therefore expects higher pollution values in some periods. For SO2 estimates, consistent intensity and relatively narrow distribution are observed at low concentration levels, reflecting the model's expectation that future SO2 levels will generally remain low. However, a few outliers in some years indicate that possible future unexpected events should also be taken into account.

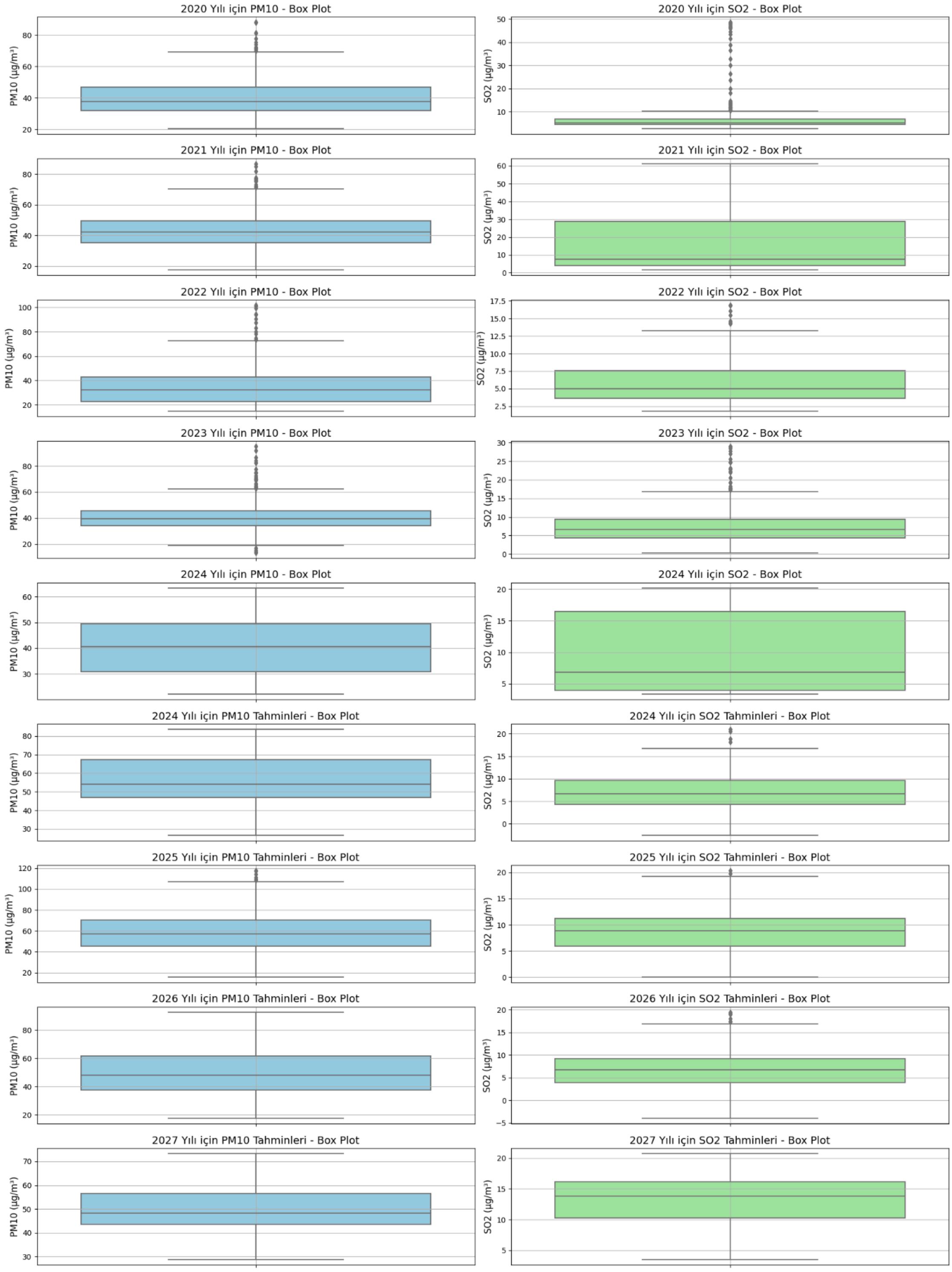

Figure 10. Visualization with Box Plot

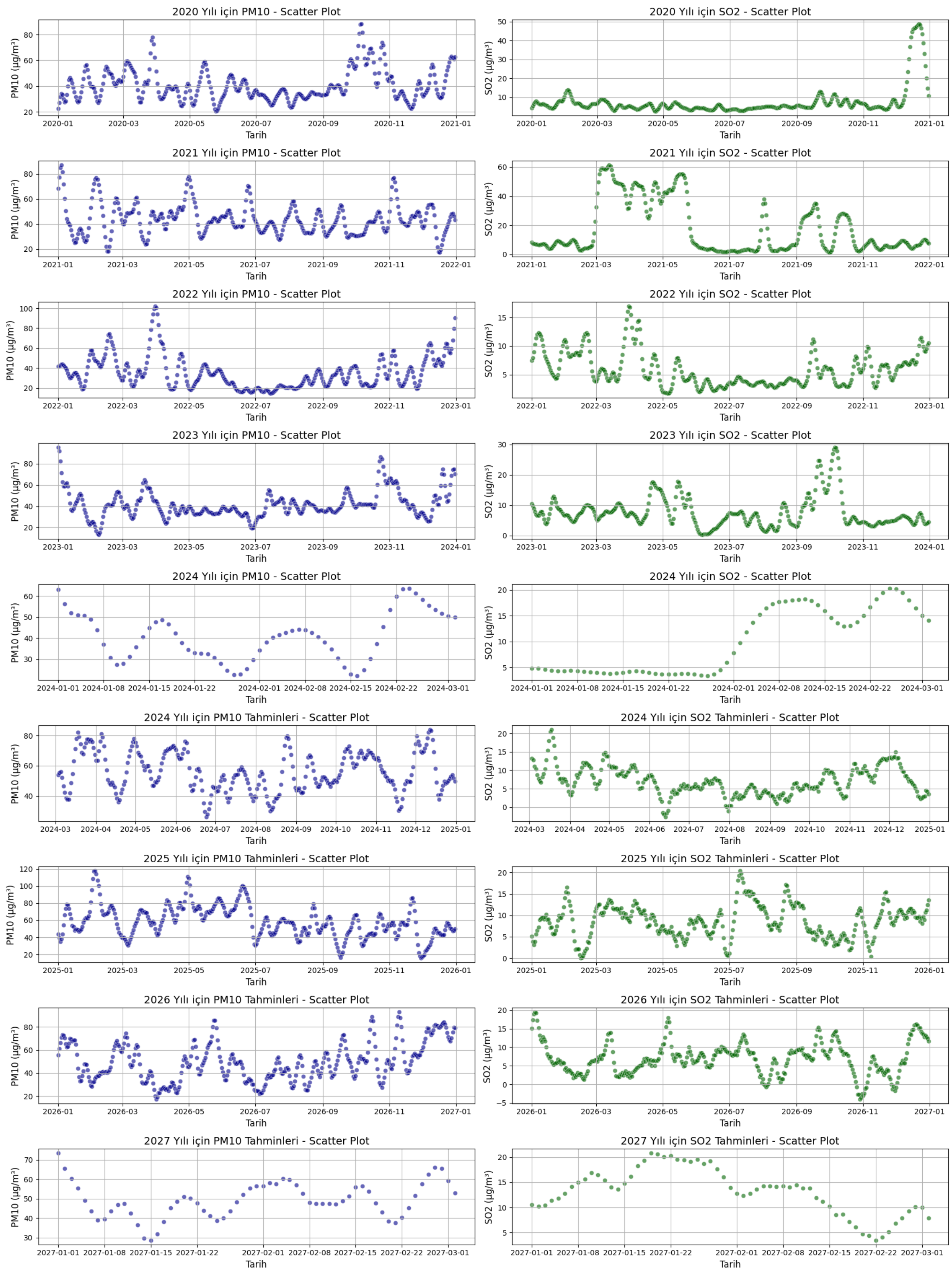

Figure 11. Visualization with Point Scatter Plot

When we look at the actual data between 2020-2024 in Figure 11, a significant fluctuation is observed in PM10 levels. There are frequent and high peaks in PM10 levels, especially in 2021 and 2023; this shows that air pollution events were more frequent and intense in these years. In 2022 and 2024, it is understood that PM10 levels were more stable and at lower levels, and sudden increases were less common. This shows that air quality was relatively better in these years. SO2 levels, on the other hand, follow a more regular course over the years and generally remain at low concentrations. However, some sudden increases were observed in SO2 levels in 2021, which shows that this year was more variable in terms of air quality. It is seen that SO2 levels were lower and stable in 2022 and 2023, and some fluctuations occurred again in 2024. When the future estimates in Figure 11 are examined, significant fluctuations and high peaks are predicted in PM10 levels, especially in 2025 and 2026. This shows that the model predicts serious changes in PM10 levels and potential pollution events in these years. For 2027, PM10 levels are expected to be lower and less variable, reflecting a better air quality expectation for this year. SO2 estimates generally remain at low levels and do not show major fluctuations. Some increases in SO2 levels are expected in 2025 and 2026, but these levels generally remain at low concentrations. SO2 levels are predicted to be quite stable in 2027. In general, these scatter plot graphics are useful in visualizing changes in PM10 and SO2 levels over time and possible future trends. Despite some uncertainties in the estimates, it can be concluded that the model successfully captures seasonal and yearly variations in air quality.

The results of this study highlight both the strengths and weaknesses of deep learning models in air quality prediction. While the RNN model demonstrated superior performance in capturing short- and long-term dependencies, its predictive accuracy in future forecasts, particularly for PM10, raises questions about the limitations of relying solely on historical data for complex, multi-faceted environmental phenomena. The relatively flat future predictions indicate that the model may struggle to represent extreme events or sudden changes, which are crucial for proactive air quality management. This underscores the importance of integrating external factors such as meteorological variability, policy interventions, or socio-economic changes to improve the robustness of predictions. Furthermore, the comparative performance of CNN and LSTM models revealed that while these architectures have potential, their limitations in handling time-series need to be addressed, perhaps through hybrid or ensemble approaches.

The visualization techniques employed, including scatter plots, violin plots, and box plots, played a significant role in interpreting and communicating the results. These tools provided an intuitive understanding of air quality patterns, seasonal variations, and model discrepancies, making the findings accessible even to non-expert audiences. However, static visualizations inherently limit real-time applications, highlighting the potential of integrating dynamic, interactive visualization systems for monitoring and forecasting. These graphical insights also revealed areas where models underperformed, such as higher variance in PM10 predictions, emphasizing the need for tailored visual analytics to complement predictive models. Addressing these challenges through adaptive modeling techniques, richer datasets, and advanced visualization frameworks could significantly enhance the applicability of such studies, ensuring that the tools developed are not only accurate but also actionable for air quality management and policymaking.

5. Conclusion

Air pollution remains a pressing environmental and health issue for individuals living in densely populated urban areas and industrialized regions. It contributes to respiratory and cardiovascular diseases, and various forms of cancer, and has broader ecological impacts, including its role in climate change through ecosystem damage. This underscores the importance of continuous monitoring and accurate prediction of air quality to mitigate its harmful effects. In this study, the air quality of the Başakşehir district in Istanbul was evaluated, and future air quality levels were predicted using deep learning methods: CNN, RNN, and LSTM. A 10-year dataset (2014–2024) was utilized to compare the performances of these models based on metrics such as Root Mean Square Error (RMSE), R^2 scores, and Mean Absolute Error (MAE). Among the models, the RNN demonstrated the highest accuracy, with the lowest error rates and the highest R^2 scores for both PM10 and SO_2 predictions. Its ability to learn both short-term and long-term dependencies in time series data made it the most effective model for predicting air pollutant levels in the near future. The LSTM model showed performance close to the RNN, especially in capturing long-term dependencies, but it fell slightly short in predictive accuracy. Meanwhile, the CNN model, while capable of capturing some patterns, struggled with the sequential and dynamic nature of the data, resulting in relatively higher error rates compared to RNN and LSTM. Using the RNN model, PM10 and SO_2 levels were forecasted for the next three years. The results revealed a reasonable level of accuracy for short-term predictions; however, the model displayed limitations in capturing complex, long-term trends. For instance, PM10 predictions showed a flatter trend with lower variance, indicating the model's difficulty in forecasting significant future changes. Similarly, while SO_2 predictions were generally concentrated at low levels, occasional variability suggested the potential for unexpected fluctuations. For longer-term predictions, hybrid models that integrate multiple methods and consider external factors may provide a more effective approach. External influences, such as meteorological variables (e.g., wind speed, temperature, precipitation), socio-economic shifts (e.g., changes in fossil fuel usage, industrial activity), and policy interventions (e.g., emission regulations, green energy incentives), play a critical role in shaping air quality over time. These factors introduce complexities that single models like RNN may not fully capture when forecasting extended periods. While hybrid models are advantageous for long-term predictions, this study highlights that single models such as RNN excel in near-future forecasts due to their ability to capture immediate temporal dependencies efficiently. Incorporating hybrid modeling approaches for long-term forecasts, alongside single models for short-term predictions, can offer a balanced and comprehensive framework for air quality prediction. Furthermore, including additional environmental and meteorological variables in future datasets would enhance predictive accuracy by accounting for the broader range of factors influencing air pollution. In conclusion, the findings demonstrate the effectiveness of deep learning models, particularly RNN, for short-term air quality predictions. Future research should focus on hybrid model development for long-term forecasting while continuing to explore ways to integrate external influences into predictive frameworks. Such advancements will ensure more reliable predictions and contribute to more effective strategies for managing and mitigating air pollution.

References

[1] R. Kaur and P. Pandey, "Air pollution, climate change, and human health in Indian cities: A brief review," Frontiers in Sustainable Cities, vol. 3, p. 705131, 2021. doi: 10.3389/frsc.2021.705131.

[2] M. B. Khan, S. Setu, N. Sultana, S. Gautam, B. A. Begum, and M. A. Salam, "Street dust in the largest urban agglomeration: Pollution characteristics, source apportionment and health risk assessment of potentially toxic trace elements," Stochastic Environmental Research and Risk Assessment, pp. 1–20, 2023. doi: 10.1007/s00477-023-02432-1.

[3] O. Isinkaralar, K. Isinkaralar, and T. N. T. Nguyen, "Spatial distribution, pollution level, and human health risk assessment of heavy metals in urban street dust at neighbourhood scale," International Journal of Biometeorology, pp. 1–13, 2024. doi: 10.1007/s00484-024-02729-y.

[4] Y. Liu, T. Jin, S. Yu, and H. Chu, "Pollution characteristics and health risks of heavy metals in road dust in Ma'anshan, China," Environmental Science and Pollution Research, vol. 30, no. 15, pp. 43726–43739, 2023. doi: 10.1007/s11356-023-25303-2.

[5] A. Lai, "Analysis of air pollution from vehicle emissions for the contiguous United States," Journal of Geovisualization and Spatial Analysis, 2022.

[6] G. D. Oreggioni, O. Mahiques, F. Monforti-Ferrario, E. Schaaf, and M. Muntean, "The impacts of technological changes and regulatory frameworks on global air pollutant emissions from the energy industry and road transport," Energy Policy, vol. 168, 2022. doi: 10.1016/j.enpol.2022.113021.

[7] T. Li, Y. Yu, and Z. Sun, "A comprehensive understanding of ambient particulate matter and its components on the adverse health effects based from epidemiological and laboratory evidence," Particle and Fibre Toxicology, vol. 19, p. 67, 2022. doi: 10.1186/s12989-022-00507-5.

[8] A. Garcia, E. Santa-Helena, and A. De Falco, "Toxicological effects of fine particulate matter (PM2.5): Health risks and associated systemic injuries—Systematic Review," Water Air and Soil Pollution, vol. 234, p. 346, 2023. doi: 10.1007/s11270-023-06278-9.

[9] A. Unnikrishnan and S. Rajeswari, "Forecasting daily air quality index and early warning system for estimating ambient air pollution on road networks using Gaussian models," Tehnički Glasnik, vol. 18, no. 4, pp. 549–559, 2024.

[10] P. Perez, C. Menares, and C. Ramírez, "PM2.5 forecasting in Coyhaique, the most polluted city in the Americas," Urban Clim., vol. 32, p. 100608, 2020, doi: 10.1016/j.uclim.2020.100608.

[11] A. Aggarwal and D. Toshniwal, "A hybrid deep learning framework for urban air quality forecasting," J. Clean. Prod., vol. 329, Art. no. 129660, 2021, doi: 10.1016/j.jclepro.2021.129660.

[12]C.-J. Huang and P.-H. Kuo, “A Deep CNN-LSTM Model for Particulate Matter (PM2.5) Forecasting in Smart Cities,” Sensors, vol. 18, no. 7, Art. no. 7, Jul. 2018, doi: 10.3390/s18072220.

[13] R. Janarthanan, P. Partheeban, K. Somasundaram, and P. Navin Elamparithi, “A deep learning approach for prediction of air quality index in a metropolitan city,” Sustainable Cities and Society, vol. 67, p. 102720, Apr. 2021, doi: 10.1016/j.scs.2021.102720.

[14] S. Masmoudi, H. Elghazel, D. Taieb, O. Yazar, and A. Kallel, “A machine-learning framework for predicting multiple air pollutants’ concentrations via multi-target regression and feature selection,” Science of The Total Environment, vol. 715, p. 136991, May 2020, doi: 10.1016/j.scitotenv.2020.136991.

[15] B. Liu, M. Lai, P. Zeng, and J. Chen, “Air pollutant prediction based on a attention mechanism model of the Yangtze River Delta region in frequent heatwaves,” *Atmospheric Research*, vol. 311, p. 107701, Dec. 2024, doi: 10.1016/j.atmosres.2024.107701.

[16] H. Yang, Z. Liu, and G. Li, “A new hybrid optimization prediction model for PM2.5 concentration considering other air pollutants and meteorological conditions,” Chemosphere, vol. 307, p. 135798, Nov. 2022, doi: 10.1016/j.chemosphere.2022.135798.

[17] P. A. Rani and Dr. V. Sampathkumar, “A novel artificial intelligence algorithm for predicting air quality by analysing the pollutant levels in air quality data in tamilnadu,” e-Prime - Advances in Electrical Engineering, Electronics and Energy, vol. 5, p. 100234, Sep. 2023, doi: 10.1016/j.prime.2023.100234.

[18] Q. Wu and H. Lin, “A novel optimal-hybrid model for daily air quality index prediction considering air pollutant factors,” Science of The Total Environment, vol. 683, pp. 808–821, Sep. 2019, doi: 10.1016/j.scitotenv.2019.05.288.

[19] B. Zhang et al., “Air pollutant diffusion trend prediction based on deep learning for targeted season—North China as an example,” Expert Systems with Applications, vol. 232, p. 120718, Dec. 2023, doi: 10.1016/j.eswa.2023.120718.

[20] J. Luo and Y. Gong, “Air pollutant prediction based on ARIMA-WOA-LSTM model,” Atmospheric Pollution Research, vol. 14, no. 6, p. 101761, Jun. 2023, doi: 10.1016/j.apr.2023.101761.

[21] G. I. Drewil and R. J. Al-Bahadili, “Air pollution prediction using LSTM deep learning and metaheuristics algorithms,” Measurement: Sensors, vol. 24, p. 100546, Dec. 2022, doi: 10.1016/j.measen.2022.100546.

[22] “Analysis of the summer thermal comfort indices in İstanbul | International Journal of Biometeorology.” Accessed: May 10, 2024. [Online]. Available: https://link.springer.com/article/10.1007/s00484-024-02669-7

[23] A. Kshirsagar and M. Shah, “Anatomization of air quality prediction using neural networks, regression and hybrid models,” Journal of Cleaner Production, vol. 369, p. 133383, Oct. 2022, doi: 10.1016/j.jclepro.2022.133383.

[24] B. Zhang et al., "Deep learning for air pollutant concentration prediction: A review," Atmospheric Environment, vol. 290, p. 119347, Dec. 2022, doi: 10.1016/j.atmosenv.2022.119347.
[25] T. D. Akinosho, M. Bilal, E. T. Hayes, A. Ajayi, A. Ahmed, and Z. Khan, "Deep learning-based multi-target regression for traffic-related air pollution forecasting," Machine Learning with Applications, vol. 12, p. 100474, Jun. 2023, doi: 10.1016/j.mlwa.2023.100474.
[26] M. Yılmaz, Y. Kara, H. C. Çulpan, G. Can, and H. Toros, "Detection and regional analysis of heatwave characteristics in İstanbul," Sustainable Cities and Society, vol. 97, p. 104789, Oct. 2023, doi: 10.1016/j.scs.2023.104789.
[27] J. González-Pardo, S. Ceballos-Santos, R. Manzanas, M. Santibáñez, and I. Fernández-Olmo, "Estimating changes in air pollutant levels due to COVID-19 lockdown measures based on a business-as-usual prediction scenario using data mining models: A case-study for urban traffic sites in Spain," Science of The Total Environment, vol. 823, p. 153786, Jun. 2022, doi: 10.1016/j.scitotenv.2022.153786.
[28] X. Shi, Z. Huang, Y. Dai, W. Du, and J. Cheng, "Evaluating emission reduction potential and co-benefits of CO2 and air pollutants from mobile sources: A case study in Shanghai, China," Resources, Conservation and Recycling, vol. 202, p. 107347, Mar. 2024, doi: 10.1016/j.resconrec.2023.107347.
[29] Y. Dokuz, A. Bozdağ, and B. Gökçek, "HAVA KALİTESİ PARAMETRELERİNİN TAHMİNİ VE MEKANSAL DAĞILIMI İÇİN MAKİNE ÖĞRENMESİ YÖNTEMLERİNİN KULLANILMASI," Ömer Halisdemir Üniversitesi Mühendislik Bilimleri Dergisi, Jan. 2020, doi: 10.28948/ngumuh.654092.
[30] S. Ünaldi and N. Yalçin, "Hava Kirliliğinin Makine Öğrenmesi Tabanlı Tahmini: Başakşehir Örneği Prediction of Air Pollution based on Machine Learning Methods: A Case Study for Başakşehir, İstanbul," p. 10.
[31] P. Aksak, Ş. K. Öztürk, and Ö. Ünsal, "Kentsel Isı Adasının İklim Parametreleri ve Uzaktan Algılama Üzerinden İncelenmesi: İstanbul Kenti Örneği," ECD, vol. 32, no. 1, Art. no. 1, Jun. 2023, doi: 10.51800/ecd.1266060.
[32] Y.-C. Lin, Y.-T. Lin, C.-R. Chen, C.-Y. Lai, and Y.-B. Wang, "Meteorological and traffic effects on air pollutants using Bayesian networks and deep learning," Journal of Environmental Sciences, Feb. 2024, doi: 10.1016/j.jes.2024.01.057.
[33] I. H. Fong, T. Li, S. Fong, R. K. Wong, and A. J. Tallón-Ballesteros, "Predicting concentration levels of air pollutants by transfer learning and recurrent neural network," Knowledge-Based Systems, vol. 192, p. 105622, Mar. 2020, doi: 10.1016/j.knosys.2020.105622.
[34] I. Jairi, S. Ben-Othman, L. Canivet, and H. Zgaya-Biau, "Enhancing air pollution prediction: A neural transfer learning approach across different air pollutants," *Environmental Technology & Innovation*, vol. 36, p. 103793, Nov. 2024, doi: 10.1016/j.eti.2024.103793.
[35] H. Dai, G. Huang, J. Wang, H. Zeng, and F. Zhou, "Prediction of Air Pollutant Concentration Based on One-Dimensional Multi-Scale CNN-LSTM Considering Spatial-Temporal Characteristics: A Case Study of Xi'an, China," Atmosphere, vol. 12, no. 12, Art. no. 12, Dec. 2021, doi: 10.3390/atmos12121626.
[36] Z. Wu, Y. Tian, M. Li, B. Wang, Y. Quan, and J. Liu, "Prediction of air pollutant concentrations based on the long short-term memory neural network," Journal of Hazardous Materials, vol. 465, p. 133099, Mar. 2024, doi: 10.1016/j.jhazmat.2023.133099.
[37] B. Das, Ö. O. Dursun, and S. Toraman, "Prediction of air pollutants for air quality using deep learning methods in a metropolitan city," Urban Climate, vol. 46, p. 101291, Dec. 2022, doi: 10.1016/j.uclim.2022.101291.
[38] "Prediction of SO2 and PM10 air pollutants using a deep learning-based recurrent neural network: Case of industrial city Sakarya | Elsevier Enhanced Reader." Accessed: Dec. 05, 2022. [Online]. Available: https://reader.elsevier.com/reader/sd/pii/S2212095521002819?token=4F8C7E774AFEF62F87DEEDFCC66D0CD0E6760E1DD3BA1655B309B6F55BDB090842B65A1488055ED705E638D6699E9847&originRegion=eu-west-1&originCreation=20221203124930
[39] J. Yang, L. Shi, J. Lee, and I. Ryu, "Spatiotemporal prediction of particulate matter concentration based on traffic and meteorological data," Transportation Research Part D: Transport and Environment, vol. 127, p. 104070, Feb. 2024, doi: 10.1016/j.trd.2024.104070.
[40] S. Jochner, I. Markevych, I. Beck, C. Traidl-Hoffmann, J. Heinrich, and A. Menzel, "The effects of short- and long-term air pollutants on plant phenology and leaf characteristics," Environmental Pollution, vol. 206, pp. 382–389, Nov. 2015, doi: 10.1016/j.envpol.2015.07.040.
[41] S. A. Ajayi, C. A. Adams, G. Dumedah, A. O. Adebanji, and W. Ackaah, "The impact of traffic mobility measures on vehicle emissions for heterogeneous traffic in Lagos City," Scientific African, vol. 21, p. e01822, Sep. 2023, doi: 10.1016/j.sciaf.2023.e01822.
[42] S. Arslankaya and Ş. Toprak, "Using Machine Learning and Deep Learning Algorithms for Stock Price Prediction," Uluslararası Muhendislik Arastirma ve Gelistirme Dergisi, vol. 13, no. 1, pp. 178–192, Jan. 2021, doi: 10.29137/umagd.771671.
[43] D. Mengüş and B. Daş, "Actionable Data Visualization for Air Quality Data in the Istanbul Location," *Balkan Journal of Electrical and Computer Engineering*, vol. 10, no. 4, Art. no. 4, Oct. 2022, doi: 10.17694/bajece.1180676.
[44] R. Rabie, M. Asghari, H. Nosrati, M. Emami Niri, and S. Karimi, "Spatially resolved air quality index prediction in megacities with a CNN-Bi-LSTM hybrid framework," *Sustainable Cities and Society*, vol. 109, p. 105537, Aug. 2024, doi: 10.1016/j.scs.2024.105537.
[45] D. Saravanan and K. Santhosh Kumar, "Improving air pollution detection accuracy and quality monitoring based on bidirectional RNN and the Internet of Things," *Materials Today: Proceedings*, vol. 81, pp. 791–796, Jan. 2023, doi:

10.1016/j.matpr.2021.04.239.
[46] G. Kurnaz and A. S. Demir, "Prediction of SO2 and PM10 air pollutants using a deep learning-based recurrent neural network: Case of industrial city Sakarya," *Urban Climate*, vol. 41, p. 101051, Jan. 2022, doi: 10.1016/j.uclim.2021.101051.
[50] T. Bernardino, M. A. Oliveira, and J. N. Silva, "Using remotely sensed data for air pollution assessment," Feb. 04, 2024, *arXiv*: arXiv:2402.06653. doi: 10.48550/arXiv.2402.06653.
[51] E. Cerezuela-Escudero, J. M. Montes-Sanchez, J. P. Dominguez-Morales, L. Duran-Lopez, and G. Jimenez-Moreno, "A systematic comparison of different machine learning models for the spatial estimation of air pollution," *Appl Intell*, vol. 53, no. 24, pp. 29604–29619, Dec. 2023, doi: 10.1007/s10489-023-05109-y.
[49] H. S. Kim *et al.*, "Development of a daily PM_{10} and $PM_{2.5}$ prediction system using a deep long short-term memory neural network model," *Atmospheric Chemistry and Physics*, vol. 19, no. 20, pp. 12935–12951, Oct. 2019, doi: 10.5194/acp-19-12935-2019.
[46] B. Cui, M. Liu, S. Li, Z. Jin, Y. Zeng, and X. Lin, "Deep learning methods for atmospheric PM2.5 prediction: A comparative study of transformer and CNN-LSTM-attention," *Atmospheric Pollution Research*, vol. 14, no. 9, p. 101833, Sep. 2023, doi: 10.1016/j.apr.2023.101833.
[47] X. Qi, G. Mei, S. Cuomo, C. Liu, and N. Xu, "Data analysis and mining of the correlations between meteorological conditions and air quality: A case study in Beijing," *Internet of Things*, vol. 14, p. 100127, Jun. 2021, doi: 10.1016/j.iot.2019.100127.
[48] D. Qu, X. Lin, K. Ren, Q. Liu, and H. Zhang, "AirExplorer: visual exploration of air quality data based on time-series querying," *J Vis*, vol. 23, no. 6, pp. 1129–1145, Dec. 2020, doi: 10.1007/s12650-020-00683-6.
[52] SIM Air Quality- Station Data Download Continuous Monitoring Center. https://sim.csb.gov.tr/STN/STN_Report/StationDataDownloadNew (202 0) (accessed 2 March 2024)
[53] A. Utku, "Hindistan'daki Turistik Şehirlerin İklim Değişkenlerinin Tahminine Yönelik Hibrit ConvGRU Modeli Hybrid ConvGRU Model for Prediction of Climate Variables of Touristic Cities in India".
[54] B. Baran, "Sıhhiye Bölgesi Hava Kalitesi İndeksinin Aşırı Öğrenme Makineleri ve Yapay Sinir Ağları ile Tahmini," 2022.
[55] Y. Dokuz, A. Bozdağ, and B. Gökçek, "HAVA KALİTESİ PARAMETRELERİNİN TAHMİNİ VE MEKANSAL DAĞILIMI İÇİN MAKİNE ÖĞRENMESİ YÖNTEMLERİNİN KULLANILMASI," *Ömer Halisdemir Üniversitesi Mühendislik Bilimleri Dergisi*, Jan. 2020, doi: 10.28948/ngumuh.654092.
[56] K. Oğuz and M. A. Pekin, "Makine Öğrenme Algoritmaları ile PM10 Konsantrasyon Tahmini," *Journal of Advanced Research in Natural and Applied Sciences*, vol. 8, no. 2, pp. 201–213, Jun. 2022, doi: 10.28979/jarnas.981202.
[57] B. Kotan and A. Erener, "PM10, SO2 hava kirleticilerinin çoklu doğrusal regresyon ve yapay sinir ağları ile sezonsal tahmini," *Geomatik*, vol. 8, no. 2, Art. no. 2, Aug. 2023, doi: 10.29128/geomatik.1158565.
[58] B. Gökçek, N. Şaşa, Y. Dokuz, and A. Bozdağ, "PM10 Parametresinin Makine Öğrenmesi Algoritmalari ile Mekânsal Analizi, Kayseri İli Örneği," *DEUFMD*, vol. 24, no. 70, Art. no. 70, Jan. 2022, doi: 10.21205/deufmd.2022247008.

Author(s) Contributions
Damla Mengus: Data collection, software, performing analysis, writing.
Bihter Das: Supervision, writing, review, and editing.

Conflict of Interest Notice
Authors declare that there is no conflict of interest regarding the publication of this paper.

Ethical Approval and Informed Consent
It is declared that during the preparation process of this study, scientific and ethical principles were followed.

Plagiarism Statement
This article has been scanned by iThenticate™.