Journal of Intelligent Systems and Internet of Things
Vol. 18, No. 02, PP. 142-156, 2026
DOI: https://doi.org/10.54216/JISIoT.180211

A Hybrid Deep Learning and Fuzzy Logic Framework for PM10 Concentration Forecasting in Istanbul

Rusul Al-bayati[1,*], Ülkü Alver Şahin[2], Hüseyin Toros[1]

[1]Istanbul Technical University Faculty of Aeronautics and Astronautics, Department of Climate Science and Meteorological Engineering Istanbul, Türkiye

[2]Department of Environmental Engineering, Engineering Faculty, Istanbul University-Cerrahpaşa, Avcılar, Istanbul, Turkey

Emails: al-bayati20@itu.edu.tr; ulkualver@iuc.edu.tr; toros@itu.edu.tr

Abstract

Air pollution, especially atmospheric particulate matter with aerodynamic diameters smaller than 10 micrometers (PM10), is one of the constant and serious environmental challenges in urban areas. Its consequences range from negative human health effects to broader ecological disruptions. With the increasing necessity of accurate and trustworthy forecasting devices in the sphere of air quality assessment, we propose a new hybrid-modeling platform that merges the sequential pattern recognition ability of Long Short Term Memory (LSTM) neural networks with fuzzy logic reasoning. The two approaches implemented in this model complement each other: while approaches taking into account the time dependence of the behavior of air pollutants address the complex temporal dynamics present in the problem, methods based on uncertainty propagate inherent uncertainties in the meteorological and environmental data. The model was trained using a well-structured, multi-variable dataset of hourly air quality and meteorological observations for five years (2019–2023) measured in Istanbul and further tested of January 2024 data. The hybrid approach outperformed all tested environments in prediction output, reaching an accuracy of 98% at the Aksaray traffic station, whereas standalone LSTM (97%) and fuzzy logic (94%) models performed lower. Importantly, it identified minute periodicity and pollution peaks with high fidelity and demonstrated robustness across diverse settings such as traffic-dense, industrial, rural and urban zones. These results place the hybrid LSTM–Fuzzy Logic model as a trusted and robust forecasting tool for predicting PM10 concentrations, providing valuable assistance to environmental policy-makers, urban planners, and public health authorities in efforts to reduce air pollution and protect the health of the population.

Received: April 09, 2025 Revised: June 12, 2025 Accepted: August 10, 2025

Keywords: Air pollution; Long Short-term Memory (LSTM); Fuzzy Logic; Hybrid mode; Machine learning; Environmental modeling

1. Introduction

Air pollution represents a significant environmental challenge that adversely affects both human health and ecosystems on a global scale. Among the various pollutants, particulate matter below 10-µm size (PM10) poses considerable risks due to its detrimental health effects, particularly in highly populated urban centers. In megacities like Istanbul, a range of factors, including anthropogenic activities and dust transport fluctuating meteorological conditions, affects the levels of PM10. Consequently, the precise forecasting of PM10 concentrations is crucial for informing environmental policies and improving air quality management practices [1].

Various predictive tools, including statistical models, machine-learning techniques, and hybrid models, have been developed over the years to enhance the accuracy of air quality forecasts. [2]The Furthermore, the effectiveness of hybrid models in enhancing prediction accuracy was assessed by contrasting their performance with that of individual models. This assessment utilized a framework that incorporated both deterministic and statistical approaches to forecast PM10 concentrations across Europe. Additionally, [3] data clustering methods were employed to refine forecast accuracy, effectively addressing the variability in PM10 levels induced by environmental factors.

Research has indicated the significance of addressing missing data and implementing effective feature selection to enhance the precision of predictive models. [4]One study examined the impact of missing data handling and feature selection on the performance of Long Short-Term Memory (LSTM) models specifically for PM10 forecasting. [5]Another investigation introduced a hybrid model that integrates fuzzy logic with time series analysis to address the variability and uncertainty inherent in air quality data, demonstrating that such hybrid approaches yield superior accuracy compared to conventional methods. [6]Additionally, time series models, including ARIMA and seasonal analysis techniques, were employed to forecast PM10 concentrations in heavily polluted urban areas in Turkey, highlighting that seasonal analysis facilitates the identification of temporal patterns, thereby improving prediction accuracy.

[7]A comprehensive examination of various interpolation methods aimed at mitigating gaps in air quality data is presented, including techniques such as averaging, linear interpolation, and those based on machine learning. The findings indicate that the implementation of advanced interpolation strategies markedly enhances the precision of assessments, underscoring the importance of effectively managing absent data within air quality forecasting models. [8]Additionally, a comparative analysis of multiple machine learning algorithms, including decision trees, support vector machines, and neural networks, reveals that these models surpass conventional approaches in detecting intricate patterns in PM10 levels, thereby facilitating more accurate air quality predictions.

Significant progress has been made in the development of air quality prediction models; however, conventional models continue to encounter difficulties in addressing non-linear temporal fluctuations and sudden shifts in PM10 concentrations. Deep learning approaches, particularly long short-term memory (LSTM) networks, have demonstrated strong capabilities in handling temporal data. Nevertheless, these models may struggle with the ambiguity and uncertainty that arise from abrupt environmental changes. Conversely, Fuzzy Logic serves as a robust method for managing ambiguity, yet it falls short in its capacity to analyze intricate temporal patterns within the data.

This study seeks to tackle the aforementioned challenges by creating a hybrid model that integrates the strengths of Long Short-Term Memory (LSTM) networks for analyzing intricate temporal patterns with the adaptability of fuzzy logic to manage uncertain data. This approach aims to enhance the precision of PM10 concentration forecasts in Istanbul. The model utilizes air quality and meteorological data gathered over a five-year span (2019-2023) from various environmental monitoring stations throughout the city. The efficacy of the hybrid model is assessed by contrasting its performance with that of discrete models, specifically LSTM and fuzzy logic, to determine its effectiveness in predicting periods characterized by abrupt fluctuations in pollution levels. Air quality forecasting faces several challenges, particularly the integration and harmonization of diverse data sources, such as meteorological records and station measurements, which may exhibit significant variability in their characteristics and accuracy. Additionally, the issue of missing data poses a substantial risk to forecast accuracy if not adequately addressed. Furthermore, managing pollution levels is complicated by a multitude of human activities and meteorological variations, especially in densely populated regions like Istanbul, where PM10 concentrations can fluctuate both temporally and spatially. Achieving accurate predictions in this context is particularly difficult when relying solely on traditional models, necessitating the development of approaches capable of accommodating ambiguous inputs and intricate temporal dynamics.

The main contribution of this study lies in the development of a novel hybrid-forecasting framework that integrates Long Short-Term Memory (LSTM) deep learning architecture with Fuzzy Logic to enhance the accuracy of PM10 concentration predictions in Istanbul. The proposed model leverages five years of hourly PM10 measurements in combination with key meteorological parameters, including temperature, relative humidity, wind speed, wind direction, rainfall, and atmospheric pressure, to ensure a comprehensive and robust predictive capability. Furthermore, the study conducts a comparative performance evaluation of three different models — LSTM, Fuzzy Logic, and the proposed Hybrid approach — highlighting the superior performance of the hybrid model. This work not only advances methodological approaches in air quality forecasting but also provides a practical framework that can be adapted for urban air quality management in other regions and for different pollutants.

2. Related works

This section deals with some of the relevant works on our research topic that have been conducted on air pollutant forecasting using Deep Learning techniques such as LSTM and hybrid approaches with Fuzzy Logic. Deep Learning models are considered more accurate than traditional techniques such as Support Vector Regression (SVR), Multiple Linear Regression (MLR), and Decision Trees (DT) when dealing with complex, nonlinear relationships in large environmental datasets.

LSTM networks have been applied to predict PM10 concentrations in urban areas using long-term hourly air quality and meteorological datasets, showing superior performance compared to ARIMA and SVR models in capturing temporal dependencies in the data [9]. A study integrated meteorological parameters such as temperature, relative humidity, wind speed, and atmospheric pressure into an LSTM-based framework for PM10 forecasting, resulting in improved prediction accuracy over models using only pollutant data [10].

Fuzzy Logic has been effectively applied to forecast PM10 concentrations by addressing uncertainty and imprecision in air quality datasets. For example, a fuzzy logic–based model was implemented to predict PM10 levels in selected urban areas of Poland, incorporating meteorological parameters such as temperature, humidity, and wind speed, and demonstrating competitive accuracy compared to conventional statistical methods [11].

Hybrid models that combine Deep Learning and Fuzzy Logic have also proven to be promising approaches for PM10 prediction. Although explicit LSTM–Fuzzy hybrid models for PM10 were not found, hybrid approaches using other combinations have shown improved forecasting performance. For instance, a hybrid model based on Multi-Objective Harris Hawks Optimization (MOHHO) and Extreme Learning Machine (ELM) achieved more stable and higher prediction accuracy for daily PM2.5 and PM10 levels compared to standalone methods[12].Another hybrid model combining LSTM with clustering methods showed enhanced accuracy in multi-station PM10 prediction tasks[13].

These studies highlight the potential of integrating Deep Learning and Fuzzy Logic to improve PM10 forecasting performance, especially when leveraging multi-year datasets and comprehensive meteorological parameters, which aligns with the objectives of the present study

3. Methodology

3.1 Study Area

Turkey is a nation characterized by its geographical diversity, positioned at the confluence of three significant bodies of water: the Aegean Sea, the Mediterranean Sea, and the Black Sea. This unique location fosters a distinct ecological and climatic environment. The country's strategic placement results in a variety of meteorological conditions that affect the dispersion of air pollutants. Istanbul, the largest city in Turkey with a population exceeding 15 million, serves as a focal point for this cultural and ecological interplay. Situated at the geographical junction of Asia and Europe, the city encounters specific challenges related to weather patterns, air quality, and urban planning. Nevertheless, it remains a vital center for culture, history, and economics. The cities geographical coordinates range from 28°10′ to 29°40′ east longitude and 40°50′ to 41°30′ north latitude [14].

Istanbul experiences a temperature range of 5-10°C during the winter months, escalating to 24°C in the summer. The winters are characterized by cold and rainy conditions, while the summers are typically hot and humid. A significant portion of the annual precipitation, totaling 677 mm, occurs between October and March. The prevailing winds in the city are predominantly from the northeast (poyraz), followed by stronger northerly winds (Yildiz) and subsequently by south-westerly winds (Ludus (.[15] The peak summer months, June, July, and August, exhibit an average high temperature of 28.7 degrees Celsius, whereas the average low during winter is recorded at 4.9 degrees Celsius, based on data spanning from 1950 to 2022. Both natural geographical factors and anthropogenic influences, such as industrial emissions, residential heating, and vehicular traffic, significantly affect the air quality in Istanbul. Additionally, meteorological conditions, including temperature inversions and wind patterns, further influence the dispersion and concentration of air pollutants [16, 17].

This research aims to examine and predict PM10 levels at various monitoring stations located in Aksaray (traffic), Arnavutköy (rural), Başakşehir-MTHM (industrial), and Kartal (urban) settings throughout Istanbul. These stations are instrumental in gathering essential data to assess pollution patterns and their relationship with meteorological conditions. Gaining insights into the spatial and temporal fluctuations of PM10 in such a densely populated and geographically intricate city is vital for the formulation of effective air quality management strategies. Figure 1 illustrates a map of the air quality monitoring stations in Istanbul utilized.

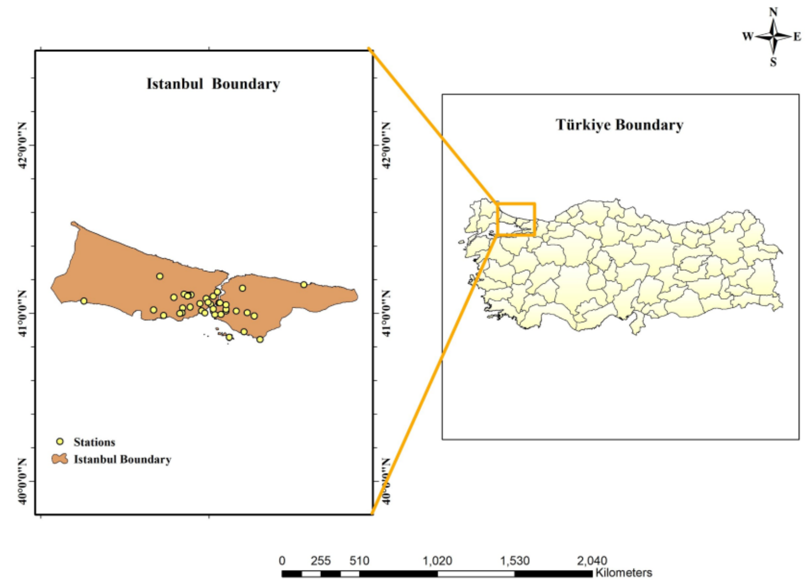

Figure 1. Location of air quality monitoring stations in Istanbul, türkiye.

3.2 Data

This research utilizes hourly PM10 concentration data obtained from four air quality-monitoring stations in Istanbul, covering the period from 2019 to 2023, as provided by the Turkish Ministry of Environment and Urbanization. The stations are located across various environments, including traffic, industrial, urban, and rural areas, thereby facilitating a thorough examination of spatial variations in air pollution. Alongside the air quality measurements, meteorological data encompassing temperature, relative humidity, wind speed, wind direction, atmospheric pressure, and precipitation were sourced from the NASA POWER database. The data coverage for these eight monitoring stations was recorded at 92.22%, indicating a high level of data integrity and continuity, with 7.78% of the data missing due to device malfunctions or transmission errors. To maintain data quality, processing was conducted using the Python programming language, which involved formatting and standardizing the data to achieve temporal alignment across different datasets, eliminating duplicate entries, and rectifying temporal discrepancies. Missing values were addressed through XGBoost interpolation techniques. The evaluation of various interpolation methods, including Moving Average and Linear Interpolation, revealed that XGBoost outperformed the others in accurately reconstructing missing values. This superiority is attributed to its capacity to leverage temporal patterns inherent in the dataset. Anomalies within the data were identified and rectified through Box Plot and Interquartile Range (IQR) analysis, where values that deviated more than three standard deviations from the mean were eliminated. This process was crucial for removing unrealistic data points that could compromise the predictive models' accuracy. Subsequently, Min-Max Scaling was applied to normalize the data, enhancing the models' training efficiency. The dataset was partitioned into a training set comprising data from 2019 to 2022, a validation set containing data from 2023, and a final test set with data from January 2024, all while preserving the chronological order to prevent data leakage between the sets during both training and evaluation.

3.3 Feature Selection and Engineering

Feature extraction and selection is a very important step in optimizing the accuracy of predictive models. This can lead to an increase in the performance of the model due to its lesser input features, thus improving accuracy, while at the same time reducing computational requirements. The Shapley Additive Explanations (SHAP) technique was used to analyze the importance of the features used in the model in order to answer the question for each variable's contribution to the model's predictions. The importance of the model parameters in predicting PM10 concentration can be observed in Figure 2, which shows the SHAP analysis indicating that wind direction (WD10M), wind speed

(WS10M), and temperature (T2M) are among the major factors contributing to PM10 centrality prediction. This is especially true for wind speed and direction, which are important drivers of the dispersion of airborne particles that shamefully affect pollution levels. Besides, temperature might affect atmospheric processes that control the buildup or scattering of pollutants. Analysis revealed that the effect of other variables, such as atmospheric pressure (PS) and precipitation rate (PRECTOTCOR), differs based on specific weather conditions.

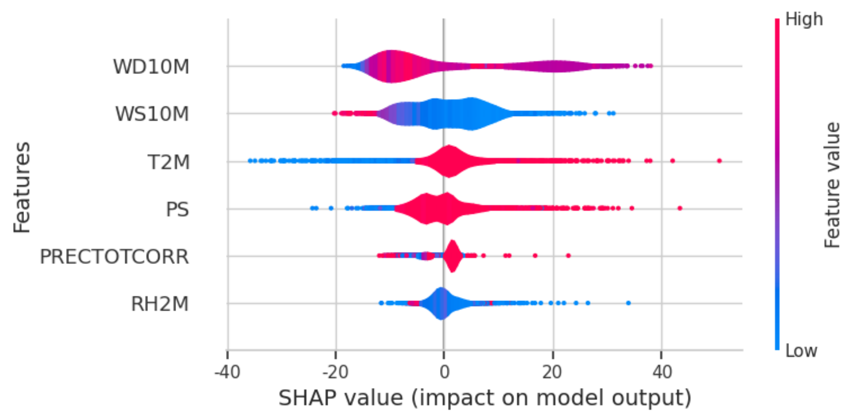

Figure 2. Analyze the importance of features using SHAP.

3.4 Algorithms and Techniques

The rapid advancement of artificial intelligence (AI) and machine learning (ML) algorithms over the past few years has placed these areas at the cutting edge of technology. Specifically, LSTM networks, Fuzzy Logic, and a novel hybrid model combining both techniques are utilized in this study, as they have shown great utility in modeling complex environmental datasets. Long Short-Term Memory (LSTM), an advanced version of Recurrent Neural Networks (RNNs), is a specialized neural network architecture designed to learn long-term dependencies in sequential data and overcome the vanishing gradient problem [18]. It is composed of three gates and a cell state, which together enable the cell to selectively retain, discard, or update information. The output gate regulates what will be passed on to the next hidden state [19]. We used a look-back period of time window to convert time-series data to a structured format applicable to supervised learning, which enhanced the prediction accuracy. Fuzzy Logic, which was first proposed by Lotfi Zadeh in 1965, adds to classical logic so that it can deal with ambiguity and uncertainty [20]. It allows for the degrees of membership between 0 and 1 (which can be represented using several mathematical functions: Gaussian, triangular, and trapezoidal shapes) unlike the classic set theory [21, 22]. Many models of human reasoning are "if-then" rules, which makes fuzzy logic especially useful in analyzing uncertain data in fields as diverse as weather forecasting and industrial controls. Moreover, advanced frameworks such as Neuro-Fuzzy Systems and Adaptive Neuro-Fuzzy Inference Systems (ANFIS) are hybrid approaches that combine fuzzy logic and artificial intelligence strategies.

Three different models were utilized in this study to predict air pollutant concentrations: LSTM, Fuzzy Logic and a Hybrid model between both approaches. We provide the same set of meteorological input variables to all models, including temperature, relative humidity, precipitation, wind speed, wind direction, and air pressure. Target outputs were PM10, PM2.5, and SO2 concentrations. Figure 3: A simplified schematic diagram representing the structure of the modeling process and the flow of information from inputs to model predictions.

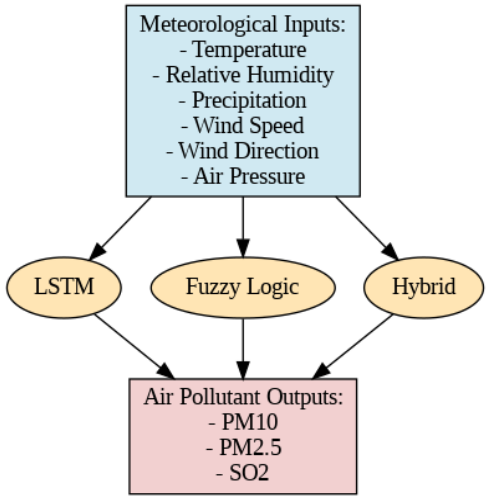

Figure 3. Framework of Air Pollution Forecasting.

3.5 Hybrid Model Architecture

The hybrid framework proposed in this study advantages long short-term memory (LSTM) networks that capture temporal patterns inherent in the climate of origin of the PM10, in conjunction with fuzzy logic, which provides a complementary structure for reducing the uncertainty in the predicted PM10 concentrations. This combined method based on deep learning for trends finding and fuzzy inference for variable dynamic evaluation enhances accuracy and reliability of sources. The method is phase wise where, climate data is passed through LSTM layers first to capture the temporal features. The generated output is further fed into a Fuzzy Inference System (FIS) that classifies the data into fuzzy sets and helps to analyze the non-linear relations between climate features and level of PM10 using the predefined fuzzy rules.

The combination of LSTM and fuzzy logic achieves good results and better accuracy in discovering different data relationships, according to new studies. We achieved this dynamic integration through fine-tuning the LSTM parameters yielding robust learning with fuzzily interpreted model rules that adapt to climate variations. The model thus combines LSTM's strength to handle sequential data and the ability of Fuzzy Logic to deal with the uncertainty to be able to act accordingly to sudden changes from the environment while still balancing the need between accuracy and interpretability Consequently, making it a powerful tool for forecasting air quality and enabling data-driven environmental policy decisions.

3.6 Evaluating and Enhancing Model Effectiveness

The efficacy of the models was systematically assessed through the application of three statistical measures: Mean Squared Error (MSE) in equation 1[23], Mean Absolute Error (MAE) in equation 2[24], and Root Mean Squared Error (RMSE) in equation 3[24].

$$MSE = \frac{\sum_{i=1}^{N} (X_i - X_m)^2}{N} \tag{1}$$

$$MAE = \sqrt{\frac{\sum_{i=1}^{n} |y_i - x_i}{n}} \tag{2}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (X_i - X_M)^2}{N}} \tag{3}$$

Finally, all three models were evaluated with a custom accuracy metric to permit comparative assessment in flexible degrees of tolerated error. The method assesses the fraction of estimates in the acceptable limit of deviation by counting all those whose predictions lay in an interval close to the real PM10 value. That is, six random tolerance thresholds were generated between 0.2 and 1.0 µg/m³. To assess performance for each threshold, the percentage of predictions within the threshold limit was computed, and then the final score was averaged across all threshold values.

The approach based on accuracy provides a useful and interpretable evaluation mechanism for the air pollution forecasting tasks. Like all classical and tolerance based performance measures were calculated only on the test set to be completely objective and not leak any information about the training or validation.

4. Results and discussion

4.1 Comprehensive PM10 Analysis

Figure 4 illustrates the temporal variations in PM10 concentration across four distinct environments in Istanbul: traffic (Aksaray), rural (Arnavutköy), industrial (Başakşehir-MTHM), and urban (Kartal). This comparison facilitates an understanding of the disparities in air quality among these locations and highlights the impact of environmental factors and anthropogenic activities on pollution levels. The traffic-monitoring site in Aksaray exhibits the most significant fluctuations in pollution, indicative of the effects of vehicular congestion [25]. In contrast, Arnavutköy, representing a rural background site, generally shows lower concentration levels, though some peaks occur—possibly due to long-range atmospheric transport of particulate matter.The industrial site in Başakşehir displays relatively stable yet elevated PM_{10} values throughout the study period, consistent with the persistent nature of emissions from industrial sources. Kartal, classified as an urban station, reveals a fluctuating pattern with relatively higher PM_{10} concentrations compared to the rural site, though not as elevated or erratic as the traffic or industrial sites. This pattern suggests the combined influence of moderate industrial activity, local traffic density, and occasional meteorological changes. Noticeable peaks in PM_{10} concentrations are evident across all stations, particularly during 2020 and 2021. These may be associated with specific meteorological events such as dust transport episodes or with shifts in human activity patterns—most notably during the COVID-19 pandemic.

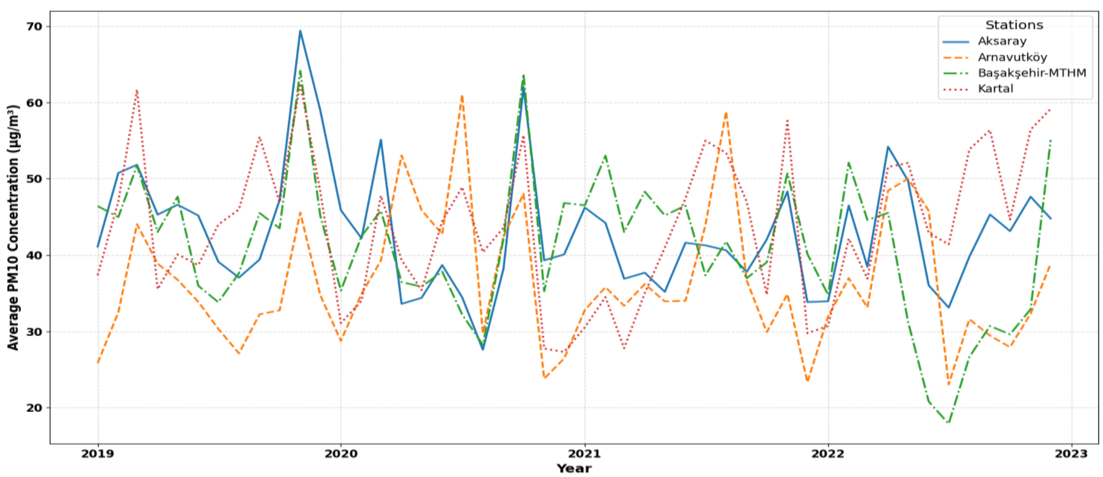

Figure 4. Monthly Average PM10 Across in Istanbul (2019-2023).

As you can see in Figure 5, shows the hourly distribution of PM10 concentrations across four monitoring stations across Istanbul, where we see differences in pollution level and high-frequency values in these times. The red dashed line represents the limit value of 50 µg/m³ as a daily average but is included purely for reference purposes as no direct comparison with hourly data is methodologically valid. However, the figure offers insight into how frequently and by how many hourly concentrations exceed this level and are, therefore, likely to contribute to violations of daily standards. The two stations with the highest hourly concentrations are the industrial station (Başakşehir-MTHM) in an industrial area with close quarters and the urban one (Kartal) with some extreme values for hourly means between 400 and 500 µg/m³ due to the density of urban architecture and heavy traffic. Traffic-related stations such as Aksaray deliver more stable pollution levels, and background stations like Arnavutköy generate lower hourly concentrations overall. Nonetheless, occasional outliers at these stations are indications of episodic pollution events, which might arise from long-range transport or local atmospheric conditions. The widespread presence of outliers at all stations reflects the high time variability of PM10 concentrations in the city.

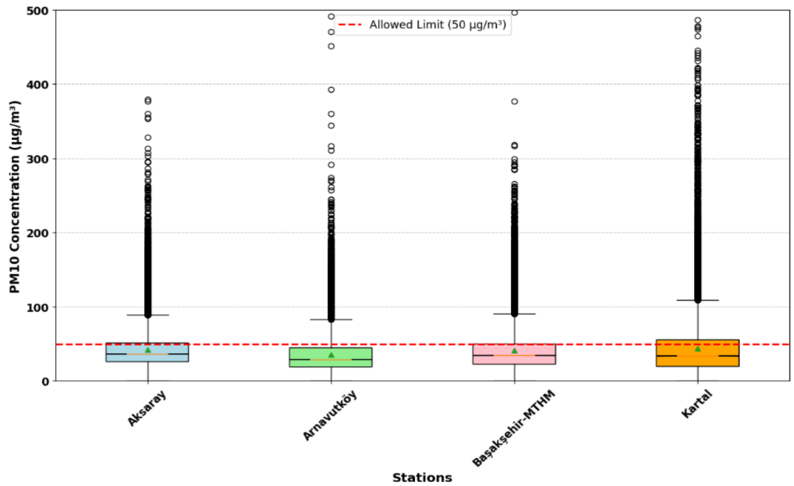

Figure 5. Box Plot of PM10 across Monitoring Stations in Istanbul (2019-2023).

4.2 Long Short-Term Memory networks (LSTM) Model

Deep learning-based models, such as LSTM, offer promising solutions for air quality prediction, outperforming traditional machine learning methods by improving feature extraction and representation. Given its ability to recognize long-term relationships in temporal data, LSTM (as a branch of deep learning) is an ideal choice for processing sequential datasets, such as air pollution measurements, enhancing the accuracy and reliability of predictions[26].Compared to conventional linear regression models, LSTM models built on deep learning outperform them, as stated in [27]. Because it can make use of the properties of time series data, LSTM successfully handles the difficulty of handling several time-dependent input variables, as pointed out in [28]. Predicting PM10 concentrations was done using the LSTM deep learning approach, which was built upon these insights.

This study represents a comprehensive evaluation of the performance of the LSTM model in predicting PM10 levels in four different environments within Istanbul: traffic, rural, urban, and industrial, using data spanning from 2019 to 2023, with a performance check using January 2024 data. The data was divided into 70% training, 20% testing, and 10% validation, allowing the model's performance to be examined at different stages and its ability to generalize. The results, as shown in Table (1) and Figures (7 and 8), reflect that the model's performance varies depending on the type of station, with the rural station (Arnavutköy) being the most accurate, while the urban station (Kartal) recorded the highest error rates. Table 1 shows the statistical values of the prediction error during the training phase, with the rural station being the most stable and having the least error, while the urban station was the least accurate due to the highly dynamic nature of air pollution in cities. These results indicate that the model was able to learn the general patterns of pollution levels during the training period, but was unable to adapt to the severe disturbances at urban and traffic stations. Figures (7and 8) show a comparison between the actual and predicted values of PM10 levels during the 2023 testing and 2024 validation phases, allowing for an assessment of the LSTM model's ability to generalize and accurately predict pollution levels. In the 2023 testing phase, which represents data not used during training, the model performed well at Aksaray (traffic) station in capturing the general trends, but faced a challenge in representing the sharp spikes caused by traffic congestion and peak times. In contrast, the model performed more accurately at Arnavutköy (rural) station, where pollution levels were less volatile and more stable, making the predictions very close to the actual values. At Kartal (urban) station, performance was less accurate, as the model was unable to adapt to constantly changing dynamic factors, such as the compounding effects of industrial activities, congestion, and rapid climate changes. At Başakşehir-MTHM (industrial) station, the model performed relatively better than at the urban station, being able to predict overall industrial pollution trends. In the validation phase (January 2024), which represents the model's ability to predict the future using data not used during training or testing, the model showed good ability to capture overall trends in PM10 levels, with performance varying between stations depending on the characteristics of each environment. At Aksaray (traffic) station, the model was able to track overall temporal patterns, but faced a challenge in capturing sudden spikes associated with traffic peaks. At Arnavutköy (rural) station, the model maintained relatively stable performance. At Kartal (urban) station, although the predictions followed the overall trend of pollution levels, sudden changes in the urban environment were more complex than the patterns learned by the model. Finally, at Başakşehir-MTHM (industrial) station, the model showed a clear improvement, being able to better predict industrial pollution trends, reflecting its adaptability to environments with regular emissions. These results indicate that the LSTM model has good ability to predict general pollution levels, with the potential to enhance its accuracy in dynamic environments by incorporating more detailed data on changing environmental factors.

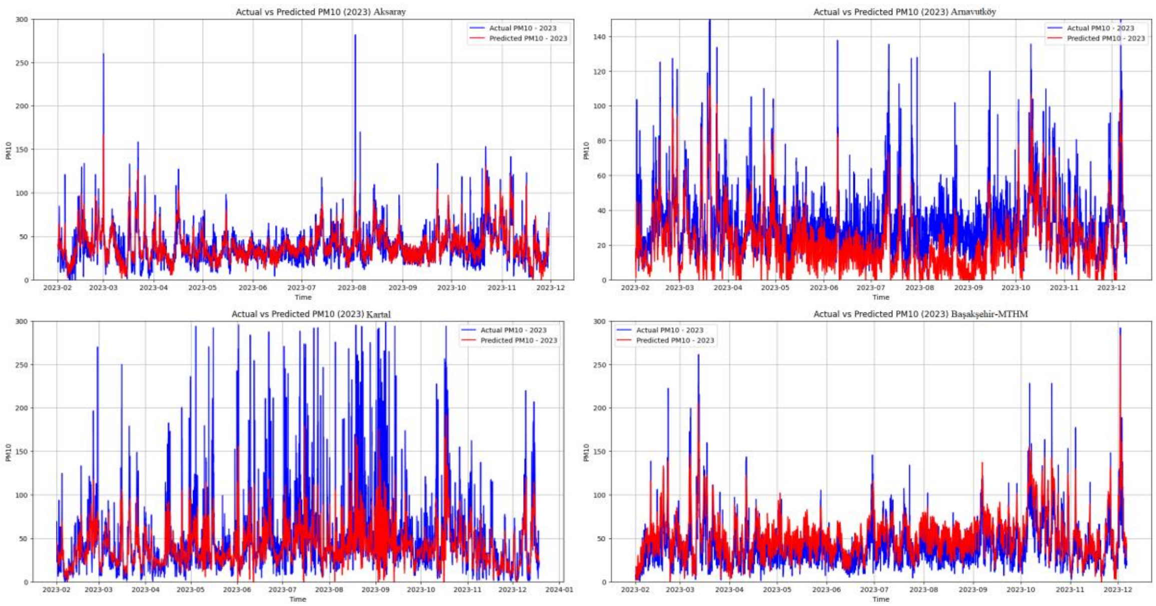

Figure 7. Comparison of observed and predicted values of PM10 levels across different stations in Istanbul for 2023.

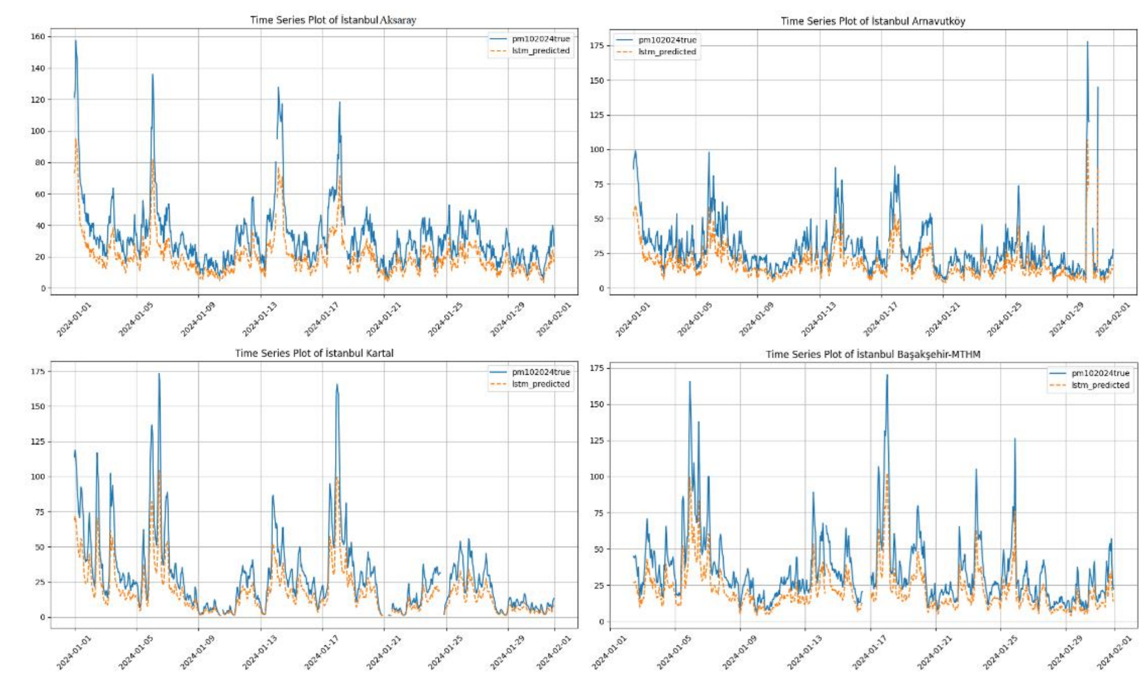

Figure 8. Comparison of observed and predicted values of PM10 levels across different stations in Istanbul for Jan. 2024.

Table 1: Performance Metrics of LSTM Model for PM10 Prediction across Istanbul Stations

Station	MSE(µg/m³)	MAE(µg/m³)	RMSE(µg/m³)
Aksaray	0.00643	0.05965	0.08017
Arnavutköy	0.00404	0.04763	0.06356
Başakşehir	0.00684	0.05696	0.08268
Kartal	0.01795	0.08601	0.13400

4.3 Fuzzy Model

Fuzzy time series forecasting with song is the second model used. Using fuzzy logic, we can deal with uncertainties and nonlinear changes in time series data in conventional FTS. There are primarily two phases to this model: fuzzification and forecasting. When fuzzification is used, the initial time series data is transformed into fuzzy values by means of pre-established fuzzy sets. We apply fuzzy rules that are based on relationships between different intervals in the time series to forecast future values in the forecasting step. The song can be utilized extensively. While using conventional FTS, we are able to successfully anticipate future values while taking into account the data's inherent uncertainties and fluctuations [29].

Analysis of the performance of the Fuzzy model in predicting PM10 levels across different stations in Istanbul, as shown in Figure 9 and Figure 10, reveals a significant variation in the accuracy of the predictions according to the type of station and its environmental characteristics. The model shows an acceptable ability to capture general trends, but suffers from weakness in accommodating sudden changes and sharp fluctuations, a common challenge in models based on fuzzy logic techniques, which tend to smooth out sharp fluctuations and produce more stable forecasts. At the Aksaray traffic station, the model performs well in simulating general pollution trends, but significantly underestimates the amplitude of peaks associated with traffic congestion. In contrast, at the Arnavutköy rural station, the agreement between actual and predicted values is high, reflecting Fuzzy's ability to predict well in environments with stable pollution, where there are no strong sudden effects as in urban and industrial areas. At the Kartal urban station, the model's performance becomes more challenging, as it fails to capture sharp fluctuations resulting from human activities, congestion, and daily climate changes. The model produces smoother values than reality, making it less accurate in urban environments, where pollution is more complex and variable. At the Başakşehir-MTHM industrial plant, performance is more stable than in urban and traffic areas, as industrial emissions often follow predictable cyclical patterns, but the model still fails to accommodate sudden changes that may result from unexpected increases in industrial emissions or climate changes that affect the spread of pollutants.

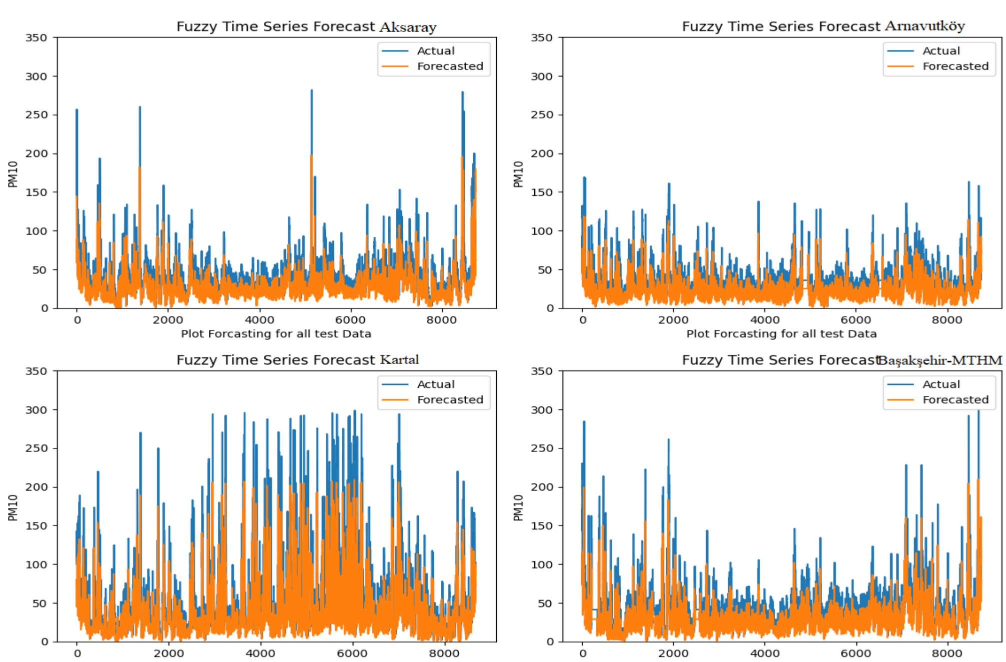

Figure 9. Performance of Fuzzy Model in Predicting PM10 Levels across Different Stations in Istanbul.

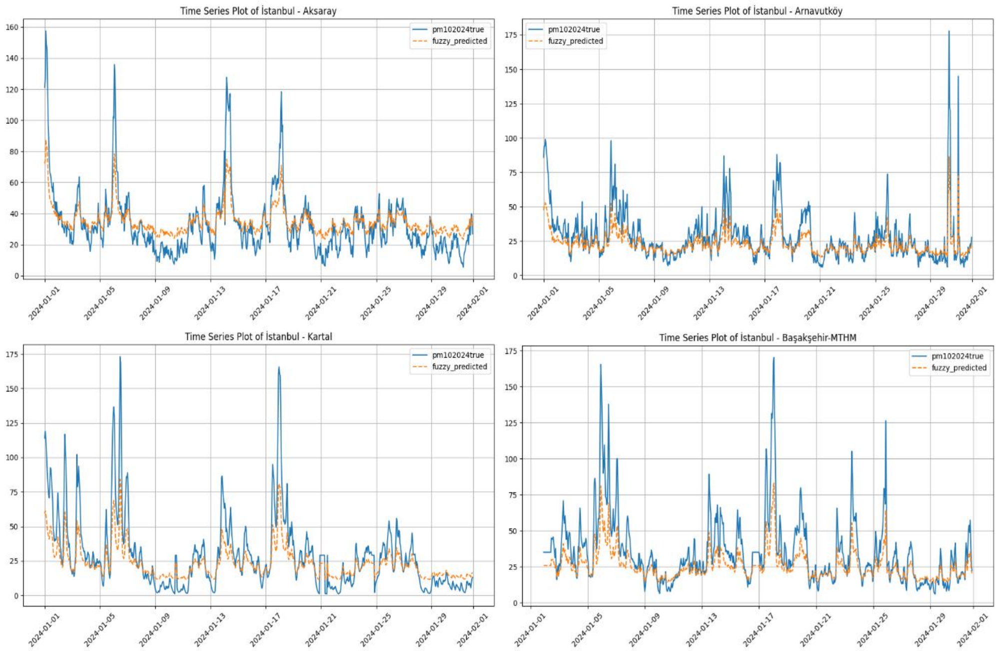

Figure 10. Evaluation of the performance of Fuzzy model in predicting PM10 levels during the test phase (January 2024) at different stations in Istanbul.

In general, the LSTM model is better at stations that are subject to dynamic and rapid changes, such as traffic and urban stations, where it can adapt to complex temporal patterns and accommodate nonlinear effects. In contrast, Fuzzy is more stable and less sensitive to sudden changes, making it more effective at rural and industrial stations where pollution levels are uniform.

4.4 Hybrid Model

In order to improve the accuracy of PM10 concentration forecasts, this study built a hybrid model that combines two prominent techniques: Long Short-Term Memory (LSTM) and Fuzzy Logic. This combined method takes advantage of both LSTM's capacity to understand past data and comprehend long-term temporal patterns and Fuzzy Logic's prowess in dealing with ambiguity and uncertainty via rule-based analysis.

From the figures (11, 12) predictions of PM10 levels using the hybrid model, which combines LSTM and fuzzy Logic, have shown a considerable improvement above those using either technique alone. The weak points of both the LSTM and Fuzzy Logic models are mitigated by the hybrid model, which combines their strengths in sequential learning and uncertainty management. The hybrid model demonstrates better balanced prediction performance in comparison to the LSTM model, which is great at capturing temporal relationships but has trouble with extreme values and spikes. When it comes to industrial and traffic-related stations, where LSTM alone would fail to either detect or fall behind fast changes, the hybrid model greatly improves accuracy during peak pollution occasions. Similarly, when compared to the fuzzy logic model, which is efficient in dealing with stable and less volatile environments such as rural stations but falters in complex and highly variable conditions, the hybrid model shows superior adaptability. The hybrid model not only maintains high performance in stable environments but also improves predictive accuracy in dynamic environments such as urban and traffic-congested stations, where fuzzy logic often performs poorly.

Table 2: A comparison of from the three models (LSTM, Fuzzy Logic and Hybrid) in prediction accuracy for four different stations from Istanbul observation sites such as traffic, rural, urban and industrial environmental characteristics. It can be seen that the Hybrid model's accuracy is greater than that obtained by LSTM and Fuzzy at all stations. This confirms the capability of the model to merge strengths from both deep learning and rule-based reasoning. This analysis demonstrates that the Hybrid model outperforms the other models, specifically in environments with rapidly changing dynamics, such as cities and traffic-congested places. The LSTM model was found to be good in terms of capturing temporal patterns, the Fuzzy Logic model had continuous performance in more stable environments such as industrial zones, and the Hybrid model uses a combination of the two. Finally, the Hybrid

model is suggested as the benchmark for PM10 models and forecasts, especially for urban and traffic stations, while Fuzzy Logic can still be an attractive alternative when utilized in industrial stations where their emissions do not vary substantially.

A recent study by Wang et al. (2023) proposed a Fuzzy Inference-Based LSTM (FLSTM) model for time series forecasting, which performed better than traditional models (standard LSTM and ARIMA) across several datasets such as PM2. 5 from Beijing and Shanghai. Related to this study, the proposed Hybrid (LSTM + Fuzzy) model obtained higher prediction accuracy for PM10 across all stations compared to other models. The conclusion of both papers is that combining temporal learning with uncertainty handling improves performance on tasks in complex environments [30].

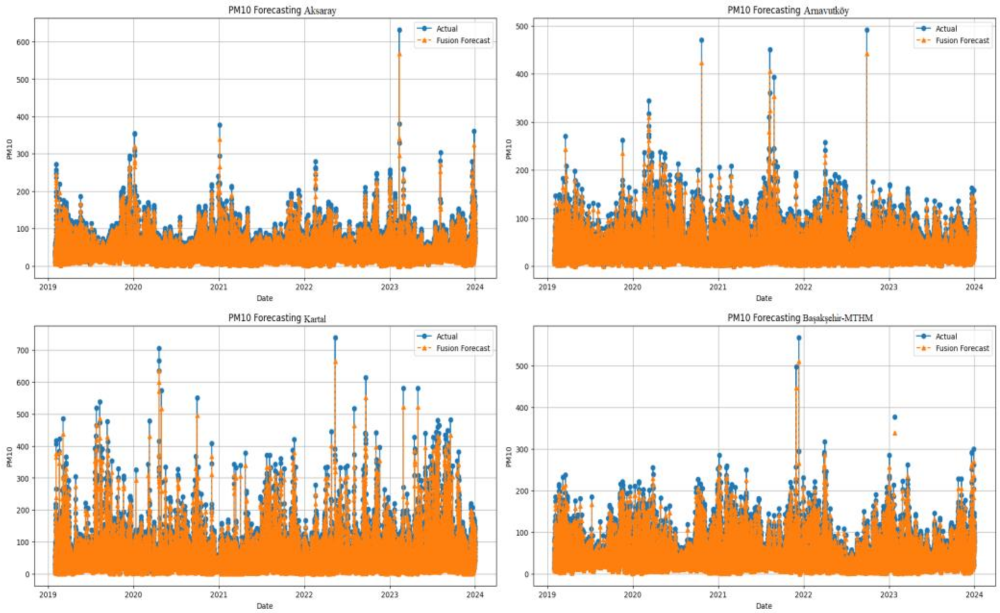

Figure 11. Performance of Hybrid Model in Predicting PM10 Levels across Different Stations in Istanbul.

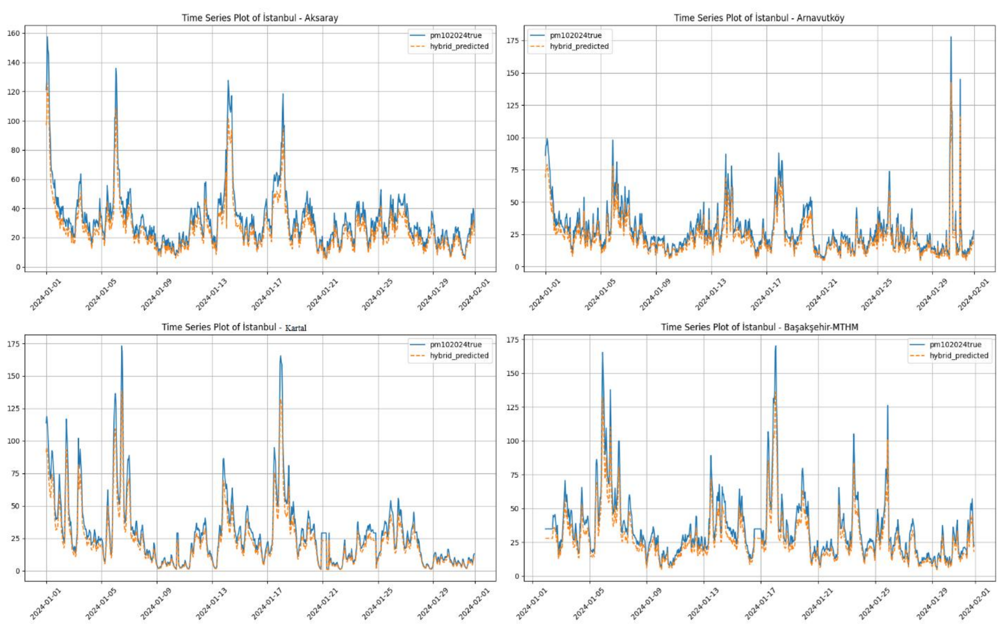

Figure 12. Evaluation of the performance of the hybrid model in predicting PM10 levels during the test phase (January 2024) at different stations in Istanbul.

Table 2: Accuracy Comparison of Prediction Models (LSTM, Fuzzy, and Hybrid) Across Different Stations.

Station	LSTM	Fuzzy	Hybrid
Aksaray	97%	94%	98%
Arnavutköy	96%	95%	98%
Başakşehir-MTHM	92%	95%	97%
Kartal	94%	92%	95%

5. Conclusion

Both the nature of the target environment and the selection of the optimal predictive model have implications on PM10 forecasting accuracy. In this study, the performance of three models (LSTM, Fuzzy and Hybrid Models) was evaluated in four different environments (traffic, rural, urban and industrial) throughout Istanbul, using data from 2019 until 2023 and validation of performance in January 2024. Hybrid yielded best accuracy and stability out of all three models resulting accuracy of 95% to 98% thus out performing both LSTM and Fuzzy even in environment like polluted urban and traffic where pollution changes extensively throughout the day. Although techniques like LSTM demonstrated challenges in capturing sharp peaks in a manner that was both responsive and accurate, while one like Fuzzy may have been stable but ultimately would lose responsiveness to sharp changes, the Hybrid provided an effective balance of accuracy and stability. The Hybrid model was particularly successful during 2023's training period and in demonstrating its ability to predict both long-term trends and short-term fluctuations during those trend instances, while LSTM and Fuzzy models performed poorly over 2024's validation period and did not show the same adaptability to unused input data. If developed further, a hybrid model predicting air quality should become a new standard in air quality prediction and a new path toward precise air quality prediction based on comprehensive data regarding environmental factors.

Acknowledgements: I would like to thank Turkish Ministry of Environment and Urbanisation providing valuable data and technical support for this research. They have been invaluable in the analysis and movement of the study.

Author Contribution; All authors contributed to the study conception and design. Data collection and analysis were performed by Rusul al-bayati. Hüseyin Toros wrote the first draft of the manuscript and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding sources: No specific funding was received for this study from any funding agency in the public, commercial, or not-for-profit sectors.

References

[1] O. W. Alqaysi and D. Z. Şeker, "PM10 air pollutant prediction using deep learning LSTM model: A case study of Istanbul, Türkiye," *Intercontinental Geoinformation Days*, vol. 7, pp. 254-257, 2023.

[2] I. Konovalov, M. Beekmann, F. Meleux, A. Dutot, and G. Foret, "Combining deterministic and statistical approaches for PM10 forecasting in Europe," *Atmospheric Environment*, vol. 43, no. 40, pp. 6425-6434, 2009.

[3] J.-M. Poggi and B. Portier, "PM10 forecasting using clusterwise regression," *Atmospheric Environment*, vol. 45, no. 38, pp. 7005-7014, 2011.

[4] İ. Aksangür, B. Eren, and C. Erden, "Evaluation of data preprocessing and feature selection process for prediction of hourly PM10 concentration using long short-term memory models," *Environmental Pollution*, vol. 311, p. 119973, 2022.

[5] Y. Alyousifi, M. Othman, A. Husin, and U. Rathnayake, "A new hybrid fuzzy time series model with an application to predict PM10 concentration," *Ecotoxicology and Environmental Safety*, vol. 227, p. 112875, 2021.

[6] H. O. Cekim, "Forecasting PM10 concentrations using time series models: A case of the most polluted cities in Turkey," *Environmental Science and Pollution Research*, vol. 27, no. 20, pp. 25612-25624, 2020.

[7] S. Ç. Bilgin, H. Toros, and T. T. Bilgin, "Evaluation of different imputation methods for handling missing air pollution data in Bursa, Turkey," *International Journal of New Horizons in the Sciences*, pp. 62-66, 2023.

[8] L. Mampitiya, N. Rathnayake, Y. Hoshino, and U. Rathnayake, "Performance of machine learning models to forecast PM10 levels," *MethodsX*, vol. 12, p. 102557, 2024.

[9] T. Xayasouk, H. Lee, and G. Lee, "Air pollution prediction using long short-term memory (LSTM) and deep autoencoder (DAE) models," *Sustainability*, vol. 12, no. 6, p. 2570, 2020.

[10] H. S. Kim et al., "Development of a daily PM10 and PM2.5 prediction system using a deep long short-term memory neural network model," *Atmospheric Chemistry and Physics*, vol. 19, no. 20, pp. 12935-12951, 2019.

[11] C. Miller, S. O'Neill, M. Rorig, and E. Alvarado, "Air-quality challenges of prescribed fire in the complex terrain and wildland urban interface surrounding Bend, Oregon," *Atmosphere*, vol. 10, no. 9, p. 515, 2019.

[12] P. Du, J. Wang, Y. Hao, T. Niu, and W. Yang, "A novel hybrid model based on multi-objective Harris hawks optimization algorithm for daily PM2.5 and PM10 forecasting," *Applied Soft Computing*, vol. 96, p. 106620, 2020.

[13] N. M. Ariff, M. A. A. Bakar, and H. Y. Lim, "Prediction of PM10 concentration in Malaysia using k-means clustering and LSTM hybrid model," *Atmosphere*, vol. 14, no. 5, p. 853, 2023.

[14] S. G. Bozkurt, L. Kuşak, and Ü. Akkemik, "Investigation of land cover (LC)/land use (LU) change affecting forest and seminatural ecosystems in Istanbul (Turkey) metropolitan area between 1990 and 2018," *Environmental Monitoring and Assessment*, vol. 195, no. 1, p. 196, 2023.

[15] M. Yılmaz, Y. Kara, H. Toros, and S. İncecik, "Analysis of the summer thermal comfort indices in İstanbul," *International Journal of Biometeorology*, vol. 68, no. 7, pp. 1327-1342, 2024.

[16] Ü. A. Şahin et al., "Assessment of ambient particulate matter and trace gases in Istanbul: Insights from long-term and multi-monitoring stations," *Atmospheric Pollution Research*, vol. 15, no. 5, p. 102089, 2024.

[17] S. Gündoğdu and T. Elbir, "Elevating hourly PM2.5 forecasting in Istanbul, Türkiye: Leveraging ERA5 reanalysis and genetic algorithms in a comparative machine learning model analysis," *Chemosphere*, vol. 364, p. 143096, 2024.

[18] S. M. Al-Selwi et al., "RNN-LSTM: From applications to modeling techniques and beyond—Systematic review," *Journal of King Saud University-Computer and Information Sciences*, p. 102068, 2024.

[19] Y.-H. Tsai, W.-C. Hung, S. Schulter, K. Sohn, M.-H. Yang, and M. Chandraker, "Learning to adapt structured output space for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7472-7481.

[20] L. A. Zadeh, "Fuzzy logic," in *Granular, Fuzzy, and Soft Computing*: Springer, 2023, pp. 19-49.

[21] A. Tomasiello, *Geometry of String Theory Compactifications*. Cambridge University Press, 2022.

[22] J. Jezewski et al., "Is abdominal fetal electrocardiography an alternative to Doppler ultrasound for FHR variability evaluation?" *Frontiers in Physiology*, vol. 8, p. 305, 2017.

[23] A. K. Sharma, D. P. Yadav, H. Garg, M. Kumar, and B. Sharma, "Bone cancer detection using feature extraction based machine learning model," *Comput. Math. Methods Med.*, vol. 2021, 2021, doi: 10.1155/2021/7433186.

[24] T. Damodaran et al., "Biological management of banana Fusarium wilt caused by Fusarium oxysporum f. sp. cubense tropical race 4 using antagonistic fungal isolate CSR-T-3 (Trichoderma reesei)," *Frontiers in Microbiology*, vol. 11, p. 595845, 2020.

[25] Ü. A. Şahin et al., "Temporal variations of atmospheric black carbon and its relation to other pollutants and meteorological factors at an urban traffic site in Istanbul," *Atmospheric Pollution Research*, vol. 11, no. 7, pp. 1051-1062, 2020.

[26] W. M. Septiawan and S. N. Endah, "Suitable recurrent neural network for air quality prediction with backpropagation through time," in *2018 2nd International Conference on Informatics and Computational Sciences (ICICoS), 2018: IEEE*, pp. 1-6.

[27] R. Navares and J. L. Aznarte, "Predicting air quality with deep learning LSTM: Towards comprehensive models," *Ecological Informatics*, vol. 55, p. 101019, 2020.

[28] L. Fu et al., "Clinical characteristics of coronavirus disease 2019 (COVID-19) in China: a systematic review and meta-analysis," *Journal of Infection*, vol. 80, no. 6, pp. 656-665, 2020.

[29] O. Orang, P. C. de Lima e Silva, and F. G. Guimarães, "Time series forecasting using fuzzy cognitive maps: a survey," *Artificial Intelligence Review*, vol. 56, no. 8, pp. 7733-7794, 2023.

[30] W. Wang, J. Shao, and H. Jumahong, "Fuzzy inference-based LSTM for long-term time series prediction," *Scientific Reports*, vol. 13, no. 1, p. 20359, 2023.