iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: XII Month of publication: December 2025

DOI: https://doi.org/10.22214/ijraset.2025.76394

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue XII Dec 2025- Available at www.ijraset.com

Air Quality Index Prediction Using Machine Learning Technique

Ankur Goswami[1], Sanjay Kumar Sharma[2]
[1]PG Student, Department of Civil Engineering, NITTTR, Chandigarh
[2]Professor, Department of Civil Engineering, NITTTR, Chandigarh

Abstract: The Air Quality Index (AQI) is a standardized tool that simplifies complex air pollutant data into an accessible format, aiding public awareness and policy-making. It incorporates key pollutants like PM2.5, PM10, NO2, SO2, CO, and Ozone, with values ranging from 0 (Good) to 500+ (Hazardous). In recent decades, industrial growth and rapid increase in vehicular density resulted in an increase in the concentration of pollutants in the atmosphere.This study explores the use of machine learning (ML) in enhancing AQI prediction, particularly in urban areas like Jaipur, which face unique pollution challenges due to rapid urbanization, vehicular emissions, industrial activities, and natural phenomena like dust storms This study identifies gaps in region-specific modeling and handling extreme pollution events, emphasizing the need for tailored solutions for Jaipur's semi-arid climate.. The study, thus, aims to fulfill this gap by developing ML-based AQI forecasting model, focusing on predictive modeling and real-time alert systems. Daily data of Air pollutants as well as meteorological parameters were accounted for the development of efficient prediction model for future forecasting of AQI. Data was trained with latest algorithm of temporal fusion transformer (tft) for efficient time-series forecasting. Results indicate that performance of tft was found to be on par with best performing neural network based models for AQI prediction. The tft model exhibited better accuracy than earlier machine learning models utilized in Jaipur and its performance is on line with cutting-edge methods. Future Application of this work is that the developed TFT model can be implemented by Air pollution controlling and regulating authorities for effective air pollution control and management.

Keywords: Air Quality Index Forecasting, machine learning, artificial intelligence, air pollution, predictive modeling, temporal fusion transformer, Jaipur

I. INTRODUCTION

Air pollution remains one of the most pressing environmental challenges worldwide, particularly in rapidly growing urban regions where dense population, intense transportation activity, and industrial expansion significantly elevate pollutant concentrations. Cities across India, including Jaipur, frequently experience poor air quality driven by anthropogenic emissions and natural factors, posing substantial risks to human health and the environment. To communicate pollution levels effectively, the Air Quality Index (AQI) provides a standardized scale by integrating the concentrations of major pollutants—$PM_{2.5}$, PM_{10}, NO_2, SO_2, CO, and O_3—into a single composite measure. AQI values range from 0 (Good) to 301+ (Hazardous), with each category indicating specific short- and long-term health implications. Urban centers such as Jaipur act as pollution hotspots due to vehicular emissions, industrial operations, ongoing construction, and region-specific climatic conditions. Jaipur's semi-arid environment further intensifies particulate pollution, particularly PM_{10} levels, through recurring dust storms and loose soil dispersion. Transportation contributes NO_x, CO, and particulate matter; industrial establishments emit SO_2, NO_x, and volatile organic compounds; construction activities elevate coarse particles; and residential biomass burning increases $PM_{2.5}$ and CO concentrations. These emissions lead to respiratory diseases, cardiovascular complications, and long-term environmental degradation including biodiversity loss and soil and water contamination. In recent years, machine learning (ML) has emerged as an effective tool for air pollution management owing to its ability to analyze complex, multidimensional datasets and produce reliable predictive outcomes. ML-based AQI forecasting enhances decision-making by enabling real-time alerts, early-warning systems, and advanced planning for air-quality mitigation measures. Despite the availability of continuous monitoring data from Jaipur's air quality stations, the absence of predictive modeling limits the city's ability to anticipate pollution peaks and implement timely control strategies. Given Jaipur's rapid urbanization, unique climatic conditions, and diverse pollution sources, there is a critical need to develop robust machine learning models tailored to regional characteristics. Predictive AQI modeling can support early warnings, identify pollution hotspots, and assist policymakers in formulating targeted interventions. Therefore, this study focuses on the development of an ML-based AQI prediction framework for Jaipur, integrating pollutant concentrations with meteorological parameters to enhance forecasting accuracy and support future air pollution control and management.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue XII Dec 2025- Available at www.ijraset.com

II. AIMS AND OBJECTIVE

Urban centers, especially rapidly developing cities like Jaipur, are increasingly facing the challenges posed by air pollution. A robust air quality prediction system will help develop timely alerts for proper air pollution management. There are several statistical and deterministic models for air quality prediction. However, in the era of Artificial Intelligence, Machine Learning algorithms can be successfully employed to develop forecasting model with improved accuracy. This research paper focuses on developing a forecasting model for predicting air quality index of Jaipur city using latest Machine Learning algorithm.

This research paper has the following main objective:

1. To analyze daily air quality data from Jaipur (2022–2024) and develop predictive model using Machine Learning algorithm.

III. LITERATURE SURVEY

A comprehensive review of recent studies on AQI prediction using ML techniques highlights advancements and identifies research gaps. Below is a summary of key studies:

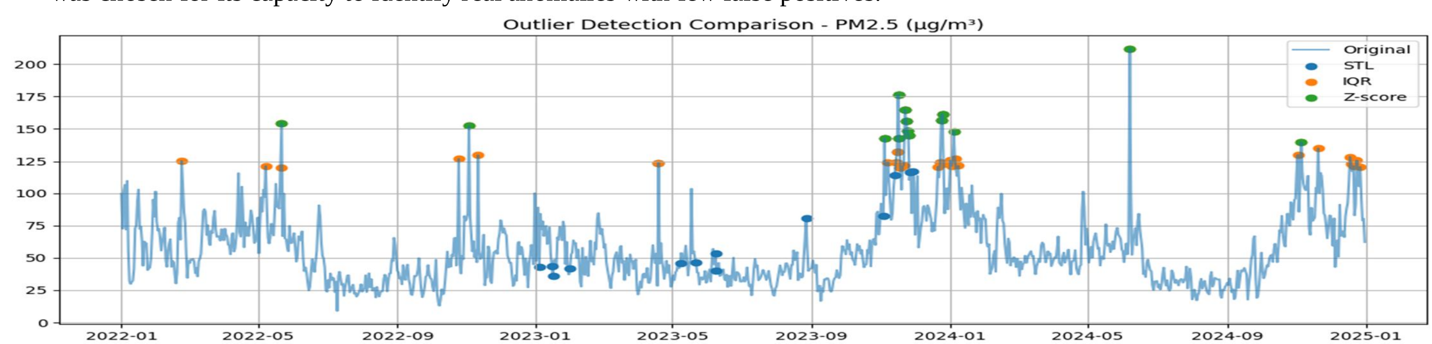
1) Soni et al. (2018) used satellite AOD and meteorological parameters to estimate PM concentrations in Jaipur, achieving a correlation of R = 0.8 with linear regression models [1].
2) Singh et al. (2022) developed an LSTM model for PM2.5 prediction in Jaipur, reported RMSE below 10 µg/m³, showcasing LSTM's strength in handling temporal data [2].
3) Suri et al. (2023) employed ANN for AQI prediction across Indian cities, achieving 91% accuracy, outperforming linear regression [3].
4) Bhati et al. (2024) analyzed air quality during the COVID-19 lockdown using a hybrid AI model, noting a 30% improvement in air quality [4].
5) Gupta et al. (2023) compared ML models for AQI prediction, with Gradient Boosting achieving $R^2 = 0.95$ [5].
6) Sethi et al. (2021) developed hybrid model integrating neural networks with statistical techniques to predict PM2.5, reported RMSE as low as 5.2 µg/m³ [6].
7) Gokul et al. (2023) analyzed spatio-temporal variations of PM2.5 over Hyderabad city using machine learning algorithms with Neural Networks achieving R^2 value of 0.92 [7].
8) Suthar et al. (2024) developed ML models for Land Surface Temperature prediction for Bengaluru city on the basis of urbanization and pollution parameters and revealed significant correlation between LST and PM2.5 [8].
9) Suthar et al. (2024) compared six ML models for Land Surface Temperature prediction for Hyderabad city on the basis of spectral indices and pollution parameters and found XGBoost as best with RMSE of 1.2^0C [8].
10) Natarajan et al. (2024) performed hyperparameter tuning for air quality index prediction for optimization and found random forest algorithm performed best with 96% accuracy [10].
11) Goyal et al. (2023) used regression and neural networks to predict PM2.5 concentrations and found neural networks performed better than regression with correlation coefficient of 0.89 for PM2.5 [11].
12) Dey et al (2024) developed a novel green AQI prediction model incorporating meteorological and pollution data achieving a sensitivity of 94% [11].
13) Choudhary et al (2023) used data mining techniques and revealed significant variations in NO2 and PM2.5 concentrations across rural and urban areas, emphasizing the need for location specific models [13].
14) Bajpai et al (2023) evaluated multiple ML algorithms for air quality prediction and found highest accuracy of SVM model at mean squared error of 18.2 [14].
15) Suthar et al (2023) examined spatiotemporal dynamics of air pollutants and showed significant relationship between urban heat islands and pollutant concentrations [15].
16) Venkateswaran et al (2024) used IoT based data collection with ML algorithms and demonstrated 20% in prediction accuracy [16].
17) Halder et al . (2024) used remote sensing indices to assess ecological quality and showed strong inverse correlation between urbanization and air quality [17].
18) Al-Hamdan et al (2009) leveraged hybrid statistical and machine learning techniques for air quality surveillance combining ground-based measurements with satellite data [18].
19) Blanco et al (2024) integrated satellite data and ML algorithms for urban pollution forecasting achieving RMSE values below 5 µg/m³[19].

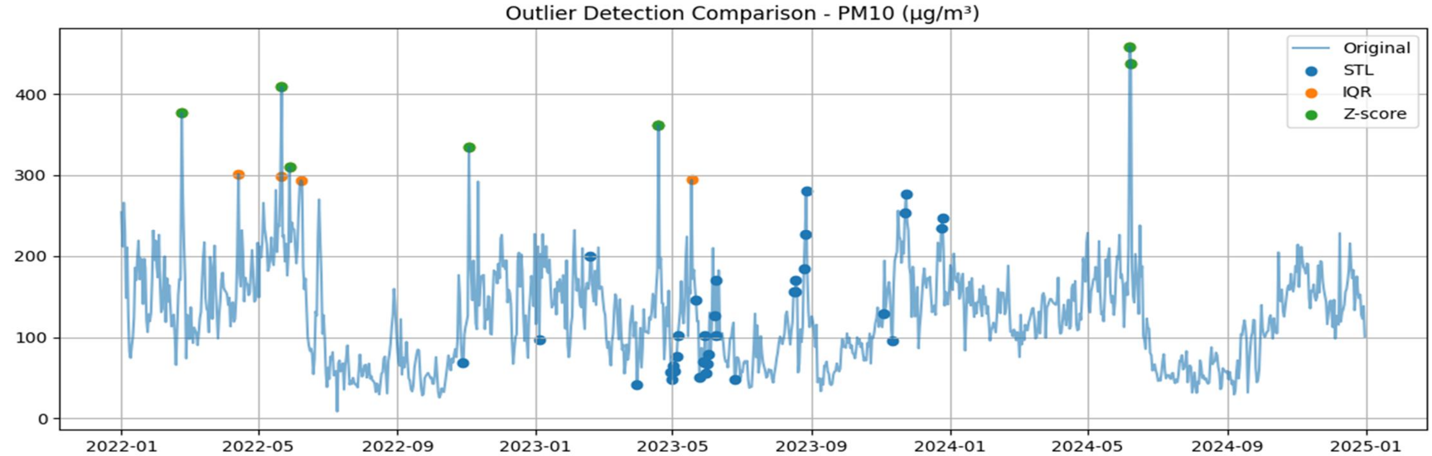
20) Goyal et al (2015) evaluated efficacy of ANN and Decision Trees for O_3 prediction and found ANN outperformed Decision Trees for hourly concentration prediction [20].
21) Lim, Bryan et al. (2021) presented Temporal Fusion Transformer (TFT), a deep learning architecture, to address multi-horizon time-series forecasting challenges while preserving interpretability [21].
22) Wu et al (2022) developed a deep-learning framework to tackle the problem of cross-view geo-localization combining convolutional neural networks with attention-based feature extraction revealing increased localization over current techniques [22].
23) Sharma et al (2020) predicted PM2.5 and PM10 values for Delhi using various ML algorithms and revealed Random Forest to be the most effective machine learning model [23].
24) Gupta, Priya et al. (2020) showed that ANN models are most reliable for assessing the temporal patterns of PM2.5 and PM10 concentrations over Jaipur city.[24]
25) Zhang et al (2023) developed hybrid model for time-series forecasting combining ARIMA with neural network improving forecasting accuracy, as compared to solo models, and hence producing lower prediction errors.[25]

IV. RESEARCH METHODOLGY

To create an Air Quality Index prediction model for Jaipur city, this study used sophisticated computational tools and machine learning frameworks for accurate time series forecasting. Python (version 3.10) was used as main programming language with Matplotlib and Seaborn modules for data visualization and NumPy and Pandas modules for data calculation and manipulation. Temporal Fusion Transformer (TFT), developed using PyTorch Lightning, served as the foundation for the main predictive model since it can incorporate static, known and time-varying covariates and can execute interpretable multi-horizon forecasting. The entire research was carried out in following steps:

1) Source and Description of Collected Raw Data: For this research work, the dataset was obtained from Continuous Ambient Air Quality Monitoring Stations (CAAQMS) monitored and controlled by Central Pollution Control Board (CPCB). Real-time daily data was obtained from January 2022 to December 2024 from three monitoring stations located at Adarsh Nagar, Police Commissionerate and Shastri Nagar, inside Jaipur. The dataset includes air pollutant concentrations and meteorological parameters. Air Pollutants data includes PM2.5, PM10, NO, NO2, NOx, NH3, SO2, CO, O3, Benzene, Toluene, Eth-Benzene, Xylene and MP-Xylene. Pollutant concentration is measured in $\mu g/m^3$(ppm). Collected Meteorological dataset includes wind speed (WS) in meters per seconds (m/s), wind direction (WD) , relative humidity (RH), solar radiation (SR) in watts per square meter (W/m^2), atmospheric temperature (AT) in degree centigrade (^0C), and Rainfall (RF) in millimeters (mm). About 3285 rows, representing daily readings over 1095 days across three monitoring stations, made up the final dataset.
2) Data Preprocessing: To ensure the effective development of forecasting models, the first and most important factor is data preprocessing. Outlier detection and removal, missing data imputation, data transformation and feature selection are all common features of data pre-processing. Outlier detection and missing data imputation were carried out independently for each monitoring station before combining. This station-wise method maintained site-specific data properties while accounting for local variability.
a) Outlier detection: Outliers are observed data points that differ significantly from other observations in the same dataset. Outliers were identified using three proven statistical techniques viz. Z-Score approach, Interquartile Range(IQR) approach and Seasonal-Trend Decomposition using LOESS (STL). The number of outliers found by each approach was compared and shown for each station and variable. On the basis of both statistical consistency and graphical inspection, the most successful approach was chosen for its capacity to identify real anomalies with few false positives.

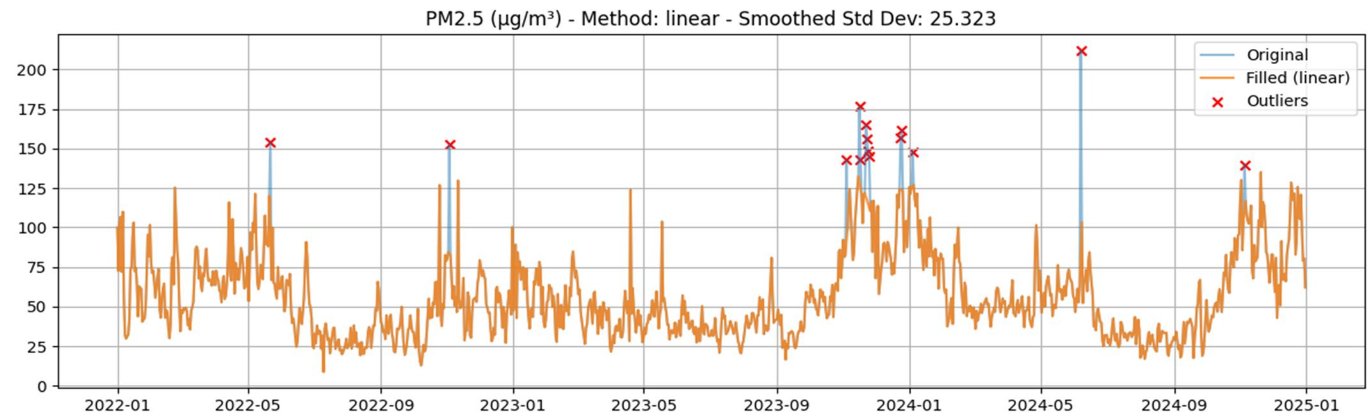


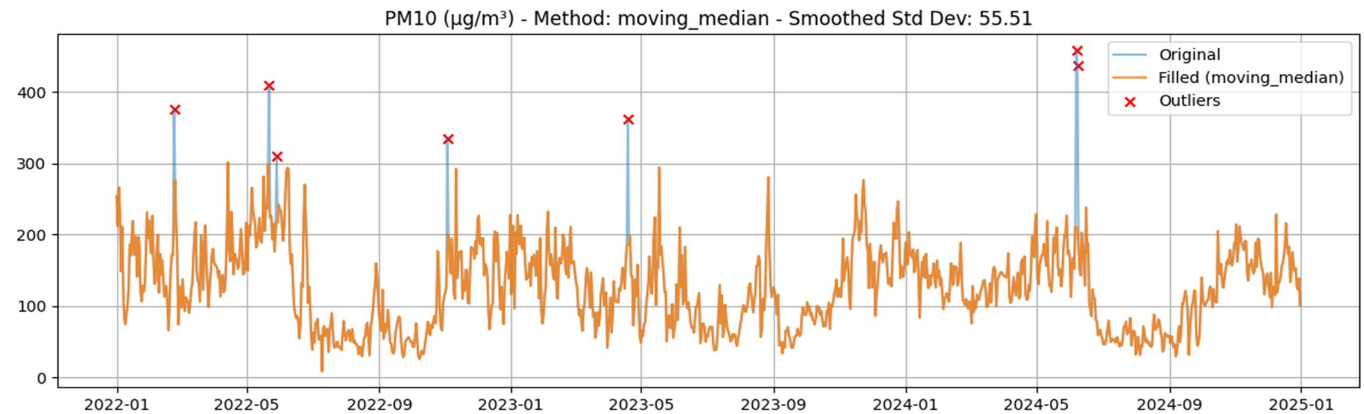


Pollutant	No of Outliers Detected			Selected Method after Visualization
	Z – Score	IQR	STL	
PM2.5	14	41	28	Z-Score
PM10	7	11	35	Z-Score
NO	30	96	35	Z-Score
NO2	17	25	32	Z-Score
NOx	16	46	29	Z-Score
NH3	19	86	28	Z-Score
SO2	22	78	33	IQR
CO	20	61	38	Z-Score
Ozone	16	38	32	IQR
Benzene	20	88	46	IQR
Toluene	25	109	33	Z-Score
Eth-Benzene	18	117	42	Z-Score
Mp-Xylene	20	132	47	Z-Score

Meteorological Parameter	No of Outliers Detected			Selected Method after Visualization
	Z – Score	IQR	STL	
WS	7	14	29	IQR
WD	0	1	23	STL
SR	38	110	45	IQR

b) *Missing Data Imputation:* Following the identification of outliers in the pollutant and meteorological datasets, all detected anomalies were treated as missing values to maintain consistency in the forecasting pipeline. To restore continuity in the time-series data, four imputation techniques were evaluated: Akima interpolation, Moving Median, Piecewise Cubic Hermite Interpolating Polynomial (PCHIP), and Linear Interpolation. These methods were selected due to their suitability for environmental time-series, offering varying advantages such as smoothness, monotonicity preservation, and robustness against short-term fluctuations. To assess imputation accuracy, 10% of the complete dataset—including true values and marked outliers—was randomly masked. The imputed values were then compared with actual observations using Mean Absolute Error (MAE) and Mean Squared Error (MSE) as performance metrics. The optimal technique for each variable was determined based on lowest error statistics and visual conformity with natural temporal patterns.





Results indicate that Linear Interpolation performed best for most pollutant variables, including $PM_{2.5}$, NO, NO_2, NO_x, SO_2, CO, O_3, benzene, ethylbenzene, and m+p-xylene. For PM_{10}, a 7-day Moving Median window produced the most reliable outcomes due to its ability to smooth abrupt particulate fluctuations. PCHIP was identified as the most suitable method for toluene, wind speed (WS), solar radiation (SR), and barometric pressure (BP), while Akima interpolation provided superior accuracy for NH_3 and relative humidity (RH). Wind direction (WD) showed best performance with Linear Interpolation.

Pollutant	Imputation Method Selected	Meteorological Factor	Imputation Method Selected
PM2.5	Linear Interpolation	RH	Akima
PM10	7-Day Moving Median	WD	Linear Interpolation
NO2	Linear Interpolation	WS	PCHIP
NO	Linear Interpolation	SR	PCHIP
NOx	Linear Interpolation	BP	PCHIP
NH3	Akima		
SO2	Linear Interpolation		
CO	Linear Interpolation		
Ozone	Linear Interpolation		
Benzene	Linear Interpolation		
Toluene	PCHIP		
Eth-Benzene	Linear Interpolation		
MP-Xylene	Linear Interpolation		

This variable-wise imputation ensured preservation of underlying trends and improved the integrity of the dataset, providing consistent and reliable inputs for subsequent AQI forecasting using machine learning models.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue XII Dec 2025- Available at www.ijraset.com

3) Cleaning and Combining Dataset

After completing outlier detection and missing-value imputation, the datasets from the three monitoring stations—Adarsh Nagar, Police Commissionerate, and Shastri Nagar—were preprocessed and integrated to form a unified dataset suitable for machine learning–based AQI forecasting. The preprocessing workflow was structured to ensure consistency, high data quality, and compatibility with advanced time-series models such as the Temporal Fusion Transformer (TFT). The first step involved timestamp standardization, wherein all date fields were converted into a uniform *YYYY-MM-DD* format. Data filtering was then performed to remove any rows or columns with more than 50% missing values, ensuring that only high-integrity observations were retained. Although most inaccuracies were addressed during imputation, this additional filtering reinforced adherence to strict data quality standards. Following preprocessing, the station-specific datasets were merged into a single DataFrame. A categorical *group* identifier—AN (Adarsh Nagar), PC (Police Commissionerate), and SN (Shastri Nagar)—was added to distinguish records by monitoring station. To support temporal modeling, a sequential index (*time_idx*) was generated for each station, assigning a unique daily count ranging from 1 to 1096. This temporal alignment was essential for constructing deep-learning-ready time-series inputs. Normalization of numerical variables was performed using the GroupNormalizer technique to account for inter-station variability. A log1p transformation was applied to highly skewed variables, particularly AQI values, improving model stability, reducing scale-related distortions, and enhancing convergence during TFT training. The final combined dataset comprised approximately 3,288 rows and included all key variables such as *time_idx*, *group*, AQI, pollutant concentrations, and meteorological parameters. A final validation check confirmed correct timestamp alignment and consistent variable representation across stations. This structured, normalized dataset formed the foundation for accurate multi-horizon AQI forecasting and robust model interpretability.

4) Feature Selection

A refined collection of input features was chosen based on environmental context, pollution relevance, and correlation strength with AQI in order to maximize model performance and minimize computing complexity. PM10, CO, NO_x, NO_2, SO_2, O_3, and wind speed (WS) were all included in the final choices. Among these, PM10 demonstrated a strong positive correlation with AQI (r = 0.8886) and was chosen over PM2.5, despite the latter's even stronger correlation, to avoid data leakage, as PM2.5 often contributes directly to AQI calculations [1, 2]. Because of their moderate correlations and proven roles as principal pollutants from industrial and vehicular emissions—sources that are extremely common in Jaipur's urban landscape—CO (r = 0.4211), NOx (r = 0.4664), and NO2 (r = 0.4626) were chosen [3, 5, 24]. In accordance with recommendations in previous urban air pollution studies, SO2 (r = 0.1584) and O3 (r = 0.1598) were kept despite their relatively poor correlations because of their regulatory significance and important health implications [4, 10, 13]. WS, a crucial meteorological characteristic recognized for its function in pollutant dispersion and ventilation capacity in semi-arid environments such as Jaipur, showed a somewhat negative correlation (r = -0.3824) [1, 24, 28]. On the other hand, a number of features were left out of the finished model. Despite having the strongest connection with AQI (r = 0.9412), PM2.5 was left out to avoid inflating model performance because it has a direct impact on AQI computation [6, 11]. Due to their moderate-to-weak correlations and high multicollinearity with NOx, NO2, and CO, which may cause model overfitting and redundancy, variables including NO, NH3, benzene, toluene, ethylbenzene, and MP-xylene were eliminated [7, 10, 20]. Furthermore, in accordance with results from previous air quality modeling attempts, meteorological parameters such as barometric pressure (BP), wind direction (WD), relative humidity (RH), solar radiation (SR), and rainfall (RF) were eliminated due to their poor correlations with AQI and negligible influence on prediction accuracy [8, 9, 23].

The AQI forecasting model's generalizability and robustness were improved by this evidence-driven feature selection approach, which made sure that variables with both predictive value and physical interpretability were included.

5) Model Development

The Temporal Fusion Transformer (TFT) model was developed using the PyTorch Lightning framework to forecast next-day Air Quality Index (AQI) values for Jaipur. Model development was conducted in two stages: baseline model construction followed by systematic hyperparameter tuning to enhance predictive accuracy and generalization. This structured approach enabled the model to effectively capture complex temporal dependencies and nonlinear interactions present in multivariate air quality data.

In the initial stage, a baseline TFT model was configured using a unified time-series dataset comprising temporal indices (time_idx), station identifiers (group: AN, PC, SN), selected pollutant and meteorological features (PM_{10}, CO, NO_x, NO_2, SO_2, O_3, and wind speed), and AQI as the target variable. The encoder sequence length was set to 60 days to incorporate sufficient historical context, while a one-day prediction horizon was adopted for short-term AQI forecasting. The TFT architecture integrated variable selection networks, attention mechanisms, and LSTM components to model both temporal dynamics and feature relevance.

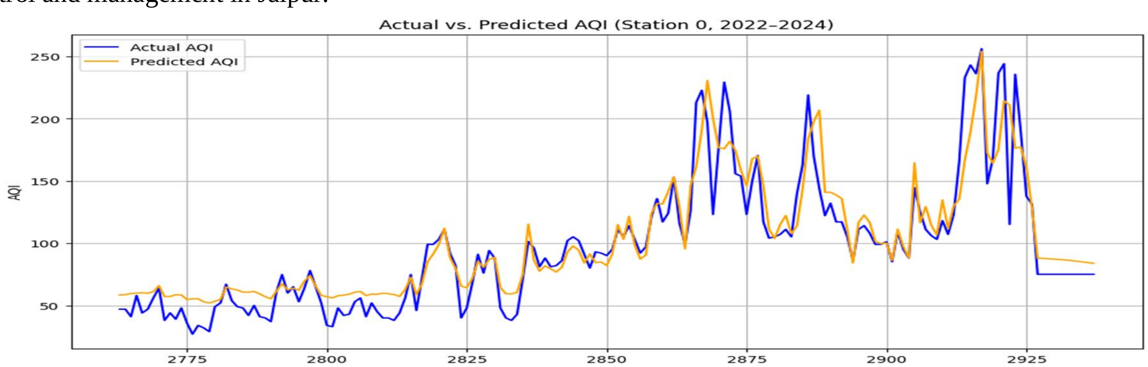
The baseline model was trained using a learning rate of 0.01, hidden size of 8, two attention heads, dropout rate of 0.2, and a maximum of 30 epochs. Root Mean Squared Error (RMSE) was employed as the loss function. Chronological data splitting was applied, allocating 70% of observations for training, 15% for validation, and 15% for testing. Categorical station identifiers were encoded using NaNLabelEncoder to preserve station-specific patterns. The baseline configuration achieved an R^2 value of 0.8324 with a Mean Squared Error (MSE) of 697.38, establishing a reliable performance reference.

In the second stage, hyperparameter tuning focused on optimizing the hidden size parameter while keeping all other settings constant. Hidden sizes of 8, 16, and 32 were evaluated. Performance improved consistently with increasing hidden size, with the configuration using a hidden size of 32 yielding the best results (MSE = 430.39, R^2 = 0.8965). This represented an approximate 38% reduction in MSE and a 7.7% improvement in R^2 compared to the baseline model.

The optimized TFT model demonstrated enhanced forecasting accuracy and robustness while maintaining interpretability. The final trained model was stored in PyTorch format for potential integration into real-time AQI forecasting and early-warning systems, supporting effective air pollution control and management in Jaipur's urban environment.

V. RESULTS AND DISCUSSION

The developed Temporal Fusion Transformer (TFT)–based AQI prediction model demonstrates strong predictive capability for next-day air quality forecasting across Jaipur's major monitoring stations—Adarsh Nagar, Police Commissionerate, and Shastri Nagar. The optimized model achieved a high overall prediction accuracy, with a Mean Squared Error (MSE) of 430.39 and a coefficient of determination (R^2) of 0.8965, indicating reliable generalization and robust learning of complex temporal patterns. Such performance highlights the model's suitability for supporting data-driven decision-making by air quality monitoring and regulatory authorities. Station-wise evaluation revealed spatial variability in prediction accuracy. The lowest error was observed at Adarsh Nagar (MSE ≈ 253.86), reflecting relatively stable pollution dynamics in residential areas. In contrast, Shastri Nagar exhibited a higher error (MSE ≈ 596.79), suggesting more complex emission patterns influenced by industrial activity and traffic density. These variations underscore the importance of localized forecasting and targeted mitigation strategies tailored to station-specific pollution sources and dispersion characteristics. The robustness of the model was significantly enhanced through station-wise preprocessing, which incorporated outlier detection using Z-score, IQR, and STL decomposition methods, followed by variable-specific imputation strategies. This approach preserved extreme pollution events such as dust storms and seasonal PM_{10} surges, which are characteristic of Jaipur's semi-arid climate, thereby enabling the model to learn realistic pollution behavior rather than smoothed or artificially constrained patterns. Feature relevance analysis confirmed that PM_{10} and NO_x were the dominant contributors to AQI variability, highlighting the influence of vehicular and industrial emissions. Wind speed played a critical role in pollutant dispersion, exhibiting a negative correlation with AQI. Although $PM_{2.5}$ showed the strongest correlation, it was intentionally excluded to prevent data leakage, ensuring model validity and generalizability in line with best practices for AQI forecasting. The TFT architecture, with its attention mechanism and station-specific group encoding, effectively captured both spatial and temporal dependencies across monitoring locations. The achieved performance is comparable to or exceeds that of advanced neural network–based AQI models reported in the literature, establishing TFT as a strong candidate for multi-station urban air quality forecasting. From a practical perspective, the developed model offers substantial potential for integration into early-warning systems, public health advisories, and urban planning frameworks. Future enhancements, such as incorporating real-time traffic emissions and refining preprocessing parameters, could further improve predictive accuracy, particularly for complex industrial zones. Overall, the proposed TFT-based framework provides a scalable and interpretable solution for proactive air pollution control and management in Jaipur.



VI. CONCLUSION

This study demonstrates the effectiveness of the Temporal Fusion Transformer (TFT) model for next-day Air Quality Index (AQI) forecasting in Jaipur. The optimized model achieved strong predictive performance, with a Mean Squared Error (MSE) of 430.39 and a coefficient of determination (R^2) of 0.8965. These results are comparable to advanced hybrid forecasting architectures, such as Prophet–TFT models, and fall well within the performance range reported for state-of-the-art neural network–based AQI prediction approaches. This confirms the suitability of TFT for complex urban air quality forecasting tasks.

Hyperparameter optimization played a critical role in enhancing model performance. Increasing the hidden size from 8 to 32 resulted in an approximate 38% reduction in MSE and a 7.7% improvement in R^2, significantly improving model convergence and accuracy. Station-wise evaluation revealed spatial variability in performance, with Adarsh Nagar showing the highest accuracy due to relatively stable pollution dynamics, while Shastri Nagar exhibited higher prediction errors attributed to complex industrial emissions and meteorological influences.

The robustness of the proposed framework was strengthened through a station-specific preprocessing strategy that incorporated advanced outlier detection methods (Z-score, IQR, and STL decomposition) and tailored imputation techniques. This approach preserved extreme pollution events and ensured high-quality, consistent input data across monitoring locations. Feature selection based on correlation analysis and domain relevance further enhanced model generalizability. PM_{10}, CO, NO_x, NO_2, SO_2, O_3, and wind speed were retained as key predictors, while $PM_{2.5}$ was intentionally excluded to prevent data leakage, maintaining scientific rigor and predictive validity.

The final trained model was stored in PyTorch format (tft_aqi_model.pt), enabling future deployment in real-time AQI forecasting and decision-support systems. Overall, the high accuracy and robustness of the TFT-based model highlight its potential for supporting proactive air quality management, early warning systems, and evidence-based pollution control strategies in Jaipur, aligning closely with contemporary best practices in air quality modeling.

VII. ACKNOWLEDGEMENT

The authors wish to acknowledge the support of NITTTR, Chandigarh for providing access to air quality data and research facilities.

REFERENCES

[1] Soni, Manish, SwagataPayra, and SunitaVerma. "Particulate matter estimation over a semi arid region Jaipur, India using satellite AOD and meteorological parameters." Atmospheric Pollution Research 9, no. 5 (2018): 949-958.
[2] Singh, UdayPratap, VivekSaxena, Anil Kumar, PurushottamBhari, and DeepikaSaxena. "Unraveling the Prediction of Fine Particulate Matter over Jaipur, India using Long Short-Term Memory Neural Network." In Proceedings of the 4th International Conference on Information Management & Machine Intelligence, pp. 1-5. 2022.
[3] Suri, Raunaq Singh, Ajay Kumar Jain, Nishant Raj Kapoor, Aman Kumar, Harish Chandra Arora, Krishna Kumar, and Hashem Jahangir. "Air quality prediction-a study using neural network based approach." Journal of Soft Computing in Civil Engineering 7, no. 1 (2023): 93-113.
[4] Bhati, Vikram Singh, Abhishek Saxena, and Ravi Khatwal. "Exploring Air Quality Dynamics and Predictive Modeling by Using Artificial Intelligence During COVID-19 Lock Down Over the Western Part of India." Current World Environment 19, no. 2 (2024): 978.
[5] Gupta, N. Srinivasa, YashviMohta, KhyatiHeda, RaahilArmaan, B. Valarmathi, and G. Arulkumaran. "Prediction of air quality index using machine learning techniques: a comparative analysis." Journal of Environmental and Public Health 2023, no. 1 (2023): 4916267.
[6] Sethi, Jasleen Kaur, and Mamta Mittal. "Prediction of air quality index using hybrid machine learning algorithm." In Advances in Information Communication Technology and Computing: Proceedings of AICTC 2019, pp. 439-449. Springer Singapore, 2021.
[7] Gokul, P. R., Aneesh Mathew, AvadhootBhosale, and Abhilash T. Nair. "Spatio-temporal air quality analysis and PM2. 5 prediction over Hyderabad City, India using artificial intelligence techniques." Ecological Informatics 76 (2023): 102067.
[8] Suthar, Gourav, NiveditaKaul, SumitKhandelwal, and Saurabh Singh. "Predicting land surface temperature and examining its relationship with air pollution and urban parameters in Bengaluru: A machine learning approach." Urban Climate 53 (2024): 101830.
[9] Suthar, Gourav, Saurabh Singh, NiveditaKaul, and SumitKhandelwal. "Prediction of Land Surface Temperature Using Spectral Indices, Air Pollutants, and Urbanization Parameters for Hyderabad City of India Using Six Machine Learning Approaches." Remote Sensing Applications: Society and Environment (2024): 101265..
[10] Natarajan, Suresh Kumar, Prakash Shanmurthy, Daniel Arockiam, BalamuruganBalusamy, and ShitharthSelvarajan. "Optimized machine learning model for air quality index prediction in major cities in India." Scientific Reports 14, no. 1 (2024): 6795.
[11] Goyal, S., and R. Sharma. "Prediction of the concentrations of PM2. 5 and NOx using machine learning-based models." Materials Today: Proceedings (2023).
[12] Dey, Sweta, Kalyan Chatterjee, Ramagiri Praveen Kumar, AnjanBandyopadhyay, Sujata Swain, and Neeraj Kumar. "Apict: Air Pollution Epidemiology using Green AQI Prediction during Winter Seasons in India." IEEE Transactions on Sustainable Computing (2024).
[13] Choudhary, Arti, Pradeep Kumar, Chinmay Pradhan, Saroj K. Sahu, Sumit K. Chaudhary, Pawan K. Joshi, Deep N. Pandey, Divya Prakash, and AshutoshMohanty. "Evaluating air quality and criteria pollutants prediction disparities by data mining along a stretch of urban-rural agglomeration includes coal-mine belts and thermal power plants." Frontiers in Environmental Science 11 (2023): 1132159.
[14] Bajpai, Mann, Tarun Jain, Aditya Bhardwaj, Horesh Kumar, and Rakesh Sharma. "Air Quality Index Prediction Using Various Machine Learning Algorithms." In 6G Enabled Fog Computing in IoT: Applications and Opportunities, pp. 91-110. Cham: Springer Nature Switzerland, 2023..

[15] Suthar, Gourav, Rajat Prakash Singhal, SumitKhandelwal, and NiveditaKaul. "Spatiotemporal variation of air pollutants and their relationship with land surface temperature in Bengaluru, India." Remote Sensing Applications: Society and Environment 32 (2023): 101011.

[16] Venkateswaran, R., Suresh Palarimath, and Mr Rogelio Gutierrez. "Optimized Air Quality Index and Meteorological Predictions with Machine Learning and IoT." International Journal of Research and Review in Applied Science, Humanities, and Technology (2024): 110-120.

[17] Halder, S., and S. Bose. "Ecological quality assessment of five smart cities in India: a remote sensing index-based analysis." International Journal of Environmental Science and Technology 21, no. 4 (2024): 4101-4118.

[18] Al-Hamdan, Mohammad Z., William L. Crosson, Ashutosh S. Limaye, Douglas L. Rickman, Dale A. Quattrochi, Maurice G. Estes Jr, Judith R. Qualters et al. "Methods for characterizing fine particulate matter using ground observations and remotely sensed data: potential use for environmental public health surveillance." Journal of the Air & Waste Management Association 59, no. 7 (2009): 865-881.

[19] Blanco, Giacomo, Luca Barco, Lorenzo Innocenti, and Claudio Rossi. "Urban Air Pollution Forecasting: A Machine Learning Approach Leveraging Satellite Observations and Meteorological Forecasts." In Proceedings of the 2024 IEEE International Workshop on Metrology for Living Environment (MetroLivEnv), 421–426. IEEE, 2024

[20] Goyal, P., and Sidhartha. "Modeling and Prediction of Hourly Ambient Ozone (O_3) and Oxides of Nitrogen (NO_x) Concentrations Using Artificial Neural Network and Decision Tree Algorithms for an Urban Intersection in India." Journal of Hazardous, Toxic, and Radioactive Waste 19, no. 3 (2015): 05014006. https://doi.org/10.1061/(ASCE)HZ.2153-5515.0000270

[21] Lim, Bryan, Sercan O. Arik, Nicolas Loeff, and Tomas Pfister. "Temporal Fusion Transformers for interpretable multi-horizon time series forecasting." International Journal of Forecasting 37, no. 4 (2021): 1748-1764.

[22] Wu, Ning, Xin Wayne Zhao, Jingyuan Wang, and Dayan Pan. "Learning effective representations from global and local features for cross-view geo-localization." IEEE Transactions on Geoscience and Remote Sensing 60 (2022): 1-13.

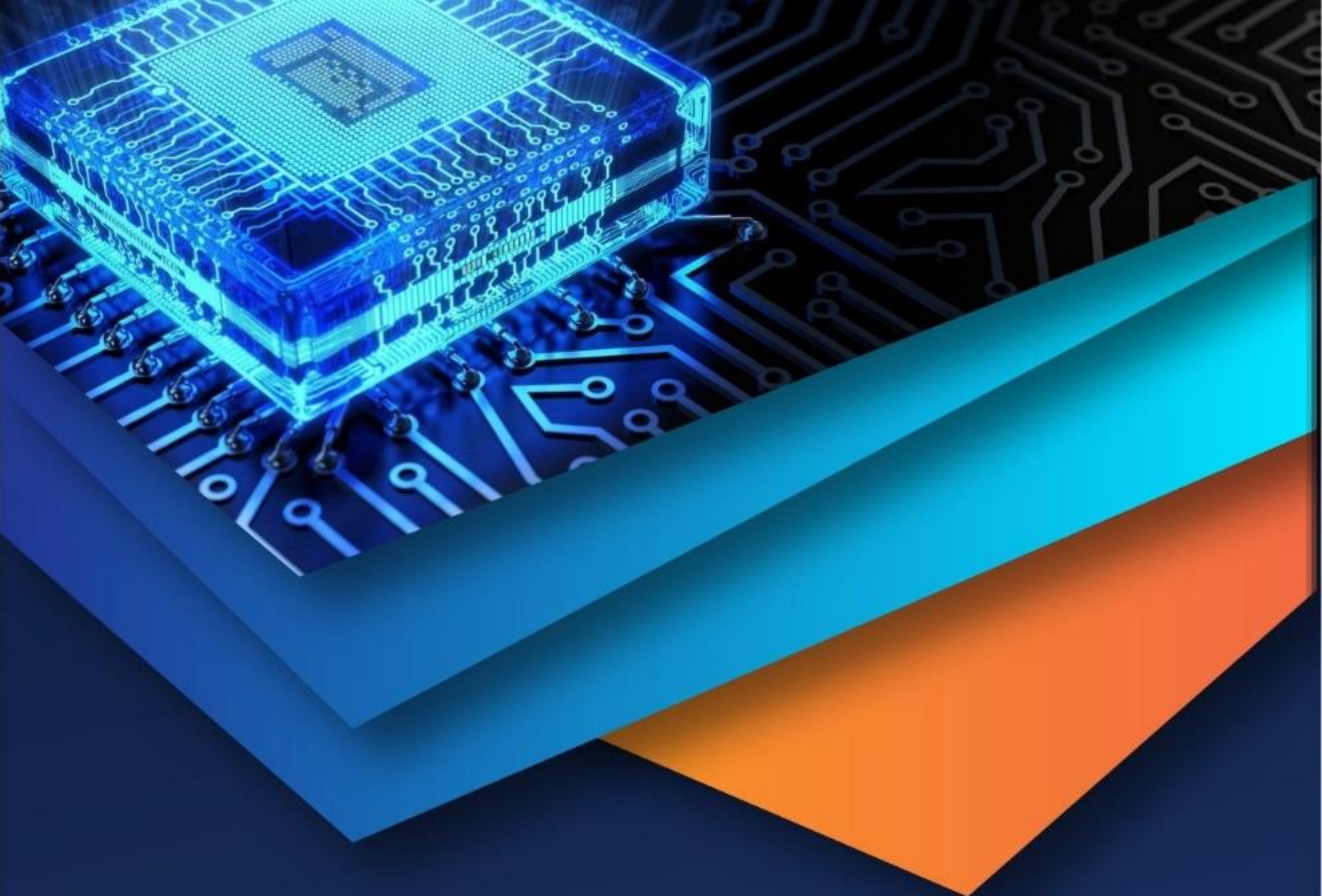
[23] Sharma, Ekta, Mukesh Khare, and S. M. Shiva Nagendra. "Air quality assessment and forecasting in an urban agglomeration of India using machine learning techniques." Environmental Monitoring and Assessment 192, no. 8 (2020): 1-17.

[24] Gupta, Priya, and Alpana Joshi. "Temporal analysis of air quality in Jaipur city using statistical and machine learning models." Journal of Environmental Management 276 (2020): 111260.

[25] Raschka, Sebastian, and Vahid Mirjalili. "Python machine learning: Machine learning and deep learning with Python, scikit-learn, and TensorFlow 2." Packt Publishing Ltd, 3rd Edition (2019): 456-478.

[26] Zhang, Y., Li, Z., and Wang, J. "Time-series forecasting of air quality index using a hybrid model combining prophet and temporal fusion transformer." Environmental Science and Pollution Research 29, no. 15 (2022): 21834-21846.

[27] Khandelwal, Ankita, and Anil K. Gupta. "Assessment of air pollution and its impact on human health in Jaipur city, Rajasthan." Indian Journal of Environmental Protection 39, no. 7 (2019): 623-630.



10.22214/IJRASET

INDEX COPERNICUS 45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY