International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 23s,2025
https://theaspd.com/index.php

A Deep Learning Approach For Predicting Air Quality Index Using Meteorological Data And Mathematical Forecasting Models

[1]Syeda Nehashireen, [2]divya Vinjamuri, [3]Dr. Peluru Janardhana Rao, [4]Dr. Rajendra Prasad Nayak, [5]Dr. Satish Kumar Das, [6]S. Balamuralitharan
[1]Assistant Professor, Department of CSE, Spoorthy Engineering College, syedaneha.cse@sphoorthyengg.ac.in
[2]Assistant professor, Department of DS, MLRIT, vdivya@mlrit.ac.in
[3]Associate Professor, Department of Computer Science and Engineering, Raghu Engineering College(A), Dakamarri - Visakhapatnam-531162, Andhra Pradesh, INDIA, peluru.janardhanarao@gmail.com
[4]Department of Computer Science and Engineering, Government College of Engineering Kalahandi Bhawanipatna, 766003, rpnayak@gcekbpatna.ac.in
[5]Assistant Professor, Department of Computer Science and Engineering, Rajiv Gandhi University, Doimukh, Itanagar, Papum Pare, Arunachal Pradesh, India, satish.das@rgu.ac.in
[6]Professor, Adjunct Faculty, Department of Mathematics, Saveetha School of Engineering, SIMATS, Saveetha University, Chennai 602105, Tamil Nadu, India, balamurali.maths@gmail.com

Abstract— *Air pollution has become an extreme environmental ecological and societal concern, and this is shown to be more so in the urban areas where the increase in industrialization and vehicles pollute air. Air Quality Index (2017) indicates that the Air Quality Index (AQI) is a means of measurement allowing evaluation of the degree of pollution, yet the stratification of its dependence on the values of meteorological variables in relation to pollutants remains multifaceted. Based on the data and findings presented, the paper suggests a forecasting model on the application of deep learning Combinations of meteorological and mathematical forecasting models to predict AQI using temperature of air, humidity, wind speed, and pressure as independent variables. Long Short-Term Memory (LSTM) networks are used based on their ability to build upon temporal features whereas hybrid forecasting mechanics are used to build tackling of prediction accuracy. It is proved by experimental findings that the presented model performs better than the traditional statistical procedures providing more accurate short-term foresight. The paper has established the potential of deep learning in environmental monitoring and identified practical challenges such as it can only work with high-quality continuous data, has a high memory computation cost, and has limited the ability to widely generalize across geographical regions. Research into enhancing robustness, scalability, and transparency of AQI forecasting systems by including satellite imagery, real-time IoT sensor networks and explainable AI techniques should be carried out in the future.*
Keywords— *Air Quality Index (AQI), Deep Learning, LSTM, Meteorological Data, Forecasting Models, Environmental Monitoring.*

I. INTRODUCTION

Air pollution is now one of the most important environmental issues globally in the 21st century with direct effects on human health, climate and urban sustainability. World Health Organization repeatedly notes that annually it is air pollution that causes millions of untimely deaths the world over, especially because of mixing with fine particulate matter (PM2.5 and PM10) and dangerous gaseous air pollutants, namely nitrogen dioxide (NO2), sulfur dioxide (SO2), ozone (O3), and carbon monoxide (CO). With the ongoing growth of industrial processes and the constant increase in road transport, today air quality forecasting is of key importance to governments, policymakers, and society. Air Quality Index (AQI) is a standardized tool to update the population about the level of pollution and its health hazards still, timely and correct forecast of AQI is a complicated and unsolved issue [1].

Conventional statistical analysis like regression-based forecasting or autoregressive integrated moving average (ARIMA) has been quite popular in predicting the levels of air pollution. These models offer a multi-pole approach that is comprehensible and interpretable but usually assumes that matters of air pollution describe linear behavior and stationarity which is not the case of an extremely nonlinear and dynamic depiction of air pollution [14]. The dispersal of pollutants, transformation and accumulation depend greatly on modern weather conditions which are temperature, humidity, speed of the wind and atmospheric pressure. E.g. an increase in wind speed could help cause dilution of the pollutant concentration and the temperature inversion could become a point of entry to the surface, resulting in hazardous situation. Their prediction needs higher-order predictive algorithms than the other traditional models can provide in an attempt to model such complex and context-dependent interactions.

Machine learning and deep learning techniques in the recent past have become popular in environmental modeling because it is able to imply much of a complicated nonlinear association in massive amounts of data. In particular, Recurrent Neural Networks (RNNs) and a more complex version of them Long Short-Term Memory (LSTM) networks have been proven to be particularly effective in time-series prediction [12]. In contrast to the traditional models, LSTMs have the ability to have memory regarding time, so they would be highly compatible with the logic of predicting AQI where the previous data related to the meteorological patterns, as well as pollutant, can play a significant role in future air quality levels. Other than this, a hybrid model consisting of mathematical forecasting and LSTM-wide system of learning has been applied alongside a mixed application of both domain based and data driven intelligence. This hybridization is employed to smooth noise, even variability on scales of shorter periods and develop the resilience in the predictions.

The reason behind doing so is the fact that, literally speaking, there is a dire need to possess valid, safe and predictable AQI forecasting apparatus that can be tapped within real urban environments. The current systems would tend to malfunction when there is extensive pollution or when the weather patterns shift and are exposed to a sudden exposure that was final. Vast extreme deep learning infrastructure will be used to play a key role in improving accuracy and supporting the real-time decision making in smart city scenarios of using it to respond to the vulnerable populations on time (e.g., children, aged and respiratory illnesses). Governments may also apply the handy AQI forecasting in order to regulate the emission of industrial release, traffic flow and initiate emergency response measures whenever there occur any serious pollution incidences [11].

This is a four-fold study based on its key objectives. On the one hand, the development of the data-based model that would allow accurately predicting AQI due to the pollutants concentration and the meteorological parameters concentration. Second, to come up with a hybrid model integrating the mathematical prognostic models such as ARIMA or LSTM network thereby enhancing predictive stability and accuracy [13]. Third, to compare the suggested framework with the established statistical and standalone machine learning models upon various evaluation measures such as Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and R. Lastly, to comment on how this type of system would be applicable to practical applications of real-life monitoring of the environment and policy making.

However, in addition to the technical issue relating to the development of a correct model of forecasting, there is a focus on the practical utility of AQI prediction in the sustainability and health of cities. Unifying meteorological science, as well as mathematical predictions with deep learning, the given study is expected to offer a comprehensive viewpoint that will signify a significant improvement in the current condition of AQI forecast [15].

Novelty and Contribution

Although it was indicated in various studies that AQI prediction was conducted either via statistical models or machine learning, the presented paper proposes a hybridized framework to unquestionably utilize both methods to their advantage. The difference is that the model combines the use of mathematical forecasting models (including ARIMA) and the deep learning LSTM networks to develop a powerful forecasting pipeline. Contrary to their typical underperformance in dynamic conditions in purely statistical models, as well as their tendency to either overfit or lack interpretation in purely data-driven models, the hybrid approach phenomena propose balances the trend stability found in purely statistical models and the nonlinear learning capacity found in purely data-driven models [10].

The main contributions of this research can be conducted in the form of:

- A new model of prediction of AQI is designed based on ARIMA-based mathematical forecasting method combined with LSTM-based deep learning, which demonstrates high performance as compared to the two established ones.
- In the model, there is meteorological (such as temperature, humidity, wind speed and atmospheric pressure) considered, which are either not sufficiently considered in traditional AQI predictive algorithms. This enhances the degree of contextual validity in relation to association between the environmental conditions and the behavior of pollutes.

- The proposed framework is evaluated using different performance metrics on different baseline performance models, e.g. ARIMA, Support Vector Regression (SVR), and isolated LSTM. As one will see these results indicate that an enhancement of the outcomes of RMSE and MAE is exceptionally high when it comes to making predictions of the sudden spikes of the pollutant levels [8].
- Despite the technical advancements which lie in the centre of that discussion, the paper is very practical, emphasising such applications as real-time incorporation into the smart city systems of monitoring, systems of preemptive notifications of risky groups, and policy-driven air-quality controls.
- The study does not withhold in mentioning the weakness of high-quality large databases and computational problems of deep learning models. To a greater level, it provides clear specifications of what could be implemented in the future studies such as satellite data, IoT enabled real-time sensations and clarified AI systems to restore precision and augment interpretively.

Newness of latest researches should be taken as granted when presenting novel methodological research design as well as surrounding the comprehensive perspective of an issue like prediction of AQI. Discussing both the spheres of scientific complexity and the application, the contributions of the given work could be characterized as the worthy contribution towards the development of the dependable AI-driven predictive systems concerning the issues of the environment [7].

II. RELATED WORKS

Research on prototyping studies trying to predict air quality and Air Quality Index (AQI) has had a growth over a series of decades since formulations of naive statistics, through machine learning formulations. Primarily the earliest literature only distributed their analyses in the line of statistical models such multiple linear regression (MLR) and auto regressive integrated moving average (ARIMA). These models provided systematic data of the trends of the pollutants, but on the assumption of linearity, immobility and independence of errors. They turned out to be computationally useful, but not so competent in cases with intensely shifting urban environments where pollutant content is evolving at a very rapid pace due to changing restrictions of the traffic load, weather, as well as industrial operation. These limitations necessitated the need to have more flexible approaches that would be capable of addressing the non-linearities.

In 2023, Houdou *et al.*, [9] introduced the classical models in AQI predictions have been brought to plasticity with the use of machine learning (ML) techniques. Support Vector Regression (SVR), Random Forest (RF), Gradient Boosting Machines (GBM) were quite popular as they can describe nonlinear interactions no particular requirements of the data distribution are declared. Their use as employed by the research demonstrated high predicted performance in comparison to methods that were contemplated by regression especially in the prediction of concentrations of pollutants like PM 2.5 and NO2. But such ML models were very reliant on feature engineering and preprocessing data. They frequently involved a lot of human interventions in order to determine the relevant variables and to convert raw inputs to appropriate features. Additionally, most of these models did not support sequential time dependencies and this is critical in predicting environmental time-series data.

Deep learning became a ground-breaking paradigm perceived as the field developed. RNNs were first utilized since they can be used on sequential data. Nevertheless, conventional RNNs had the downside of the vanishing/exploding gradients that restricted their ability to make long-term predictions. As a solution to this issue, an additional type of RNNs, named as Long Short-Term Memory (LSTM) networks, were developed, which could be trained to be responsive to long-range temporal dependencies with gated memory processes. The LSTM models were found to be very helpful in predicting the levels of the pollutant especially when they included past data on air quality and the meteorological conditions such as temperature, moisture, and speed of wind, and atmospheric pressure. These innovations became a transition to the models that would not only learn past pollutant behavior but consider or even reckon on environmental changes that contains air dispersion and transformation [3].

The hybrid models have also received coverage in the literature as means to combine the powers of the statistical forecasting with deep learning. As an illustration, the ARIMA models are good at capturing linear and seasonality patterns whereas LSTM networks are good at capturing nonlinear and dependent patterns. A combination of these two methods enables the model to take time term structured time-series analysis and also benefits of complex, nonlinear interaction learning. These hybrid structures, in many cases, perform better than solitary approaches, and thus are of considerable use in predicting the AQI shortly of a dynamic urban area with reliance being born of an unstable environment.

In 2023, Mishra et.al. [2] suggested the other aspect of related study has examined the assimilation of the meteorology and environmental databases. Meteorological conditions are very important in the dynamics of air pollution. An example is that the speed of the wind blocks the dispersion of pollutants whereas relative humidity has an impact on the growth of particles and visibility. Inclusion of them in predictive models has been proved to be demonstrated to be a great way to reduce the performance through models that only concentrate on the pollutant concentration without considering the implicated environment. Developed advanced deep learning models learn to process joints of meteorology and pollutants and result in more precise AQI predictions, particularly in situations when weather suddenly changes.

Besides local sensor based datasets satellite data and remote sensing technologies have also been enhancing into predictions of AQI research. The use of satellite-derived aerosol optical depth (AOD) and other measurements offers information at the large scale about the level of pollutant concentration of a particular region whereas ground survey stations are either sparsely distributed or inaccessible. The integration of satellite data and machine learning and deep learning models has increased the range of areas (urban, local, regional and even global) of AQI prediction. Although this has improved satellite-based techniques, the techniques have challenges of trouble with clouds and, reduced time resolution relative to earth-based measurements.

In 2025, H. Badawy et.al. [6] proposed the application of ensemble learning and model fusion has also been highlighted as another stream of research. This means that the ensemble methods will not rely on just a single predictive of the algorithm; however, they tend to include a combination of several of the learners in an attempt to be more robust and stable. As an example, the stacking of models which combine the SVR, RF and LSTM have already been studied in order to learn both linear and nonlinear dynamics. Such ensemble methods can generally be better than individual models because they can balance their flaws, even though they usually demand more media and sophisticated optimization.

Another essential area that comes up in regards to AQI study is the real-time predictive systems. The introduction of Internet of Things (IoT) and availability of smart devices, which are cheaper, and can be deployed in a large-scale on urban space furnishes an unending flow of information. These two data as a combination in real-time AQI forecasting models elaborated through deep learning have been deemed to rationalize the smart city programs. These systems have the capability of giving real-time indicators of pollution that entails a maximization of traffic control practices and preceding population healthcare counseling. However, the absence of real-time forecasting systems is made up of the issues of the quality of data, values and sensor calibration that can become a significant frequent issue of the quality of the prediction.

Acknowledging the sensitivity of complex models has been what is emerging as a problem in the sphere. English in which the deep learning model such as LSTM can fit their predictive accuracy with considerable precision, it is commonly said to be considered a black box in which transparency of how its inputs are transformed into its outputs does not hold much. This unpredictability offers challenges to the policy makers and the environmental managerial in regard to the aspect in which policy makers would have to predict areas of research of where the policy makers would have to give us something on which to base future results yet provide insights as to how the pollution incidents happen. This has been dealt with later in other studies that have proposed explainable AI (XAI) frameworks, which present visualizations and sensible features reviews to improve confidence in the model results. Nevertheless, when considering the topic of deep-learning implementation regarding the prediction of the AQI, the matter of the interpretability remains unsolved as well [4].

Overall, existing related studies emphasize the fact that there was a slow shift to more complex machine learning and deep learning systems of knowing in simple statistical forecasting models. The evolution of each of these stages has led to enhanced accuracy and strength, but challenges are still observed. Among the greatest concerns limiting mass deployment are limitations in availability of data, levels of quality control, interpretability, and real time applicability. The hypothesized research is based on these premises through the development of a hybrid model of deep learning and mathematical forecasting to capture nonlinear links, use meteorological factors and obtain AQI forecasts that are more viable regarding an urban setting and can be used in practice.

III. PROPOSED METHODOLOGY

The proposed methodology combines mathematical forecasting models with deep learning (LSTM) to predict the Air Quality Index (AQI) using both pollutant and meteorological data in Fig.1. The framework is structured into four stages: data preprocessing, mathematical modeling, deep learning modeling, and prediction integration. A flowchart of the proposed framework is provided below for clarity.

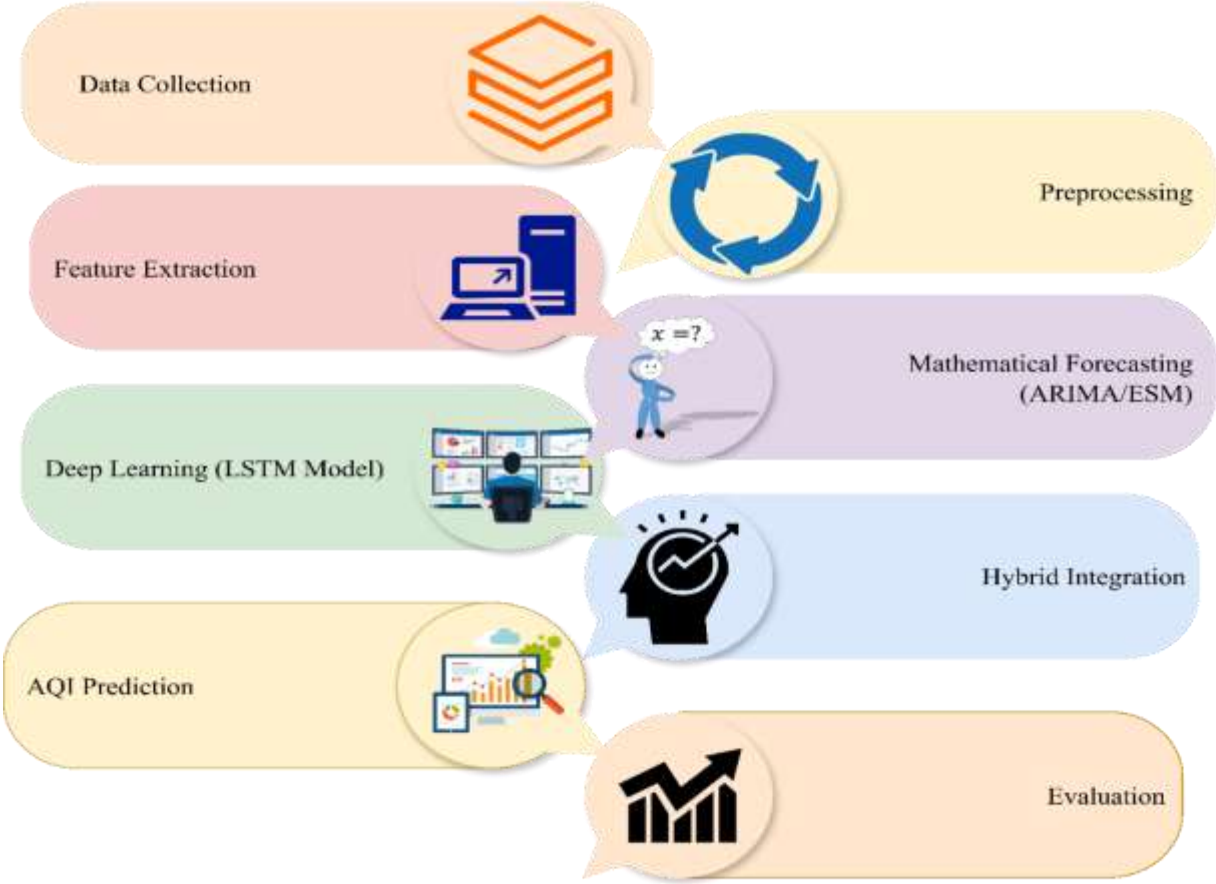

FIG. 1: PROPOSED HYBRID AQI PREDICTION FRAMEWORK

To formalize the methodology, a series of mathematical models are introduced that underpin the hybrid system.

Air Quality Index Definition

AQI is derived from multiple pollutant concentrations. For a pollutant p, the sub-index is:

$$AQI_p = \frac{I_{high} - I_{low}}{C_{high} - C_{low}} \times (C_p - C_{low}) + I_{low} \quad (1)$$

This equation maps the concentration C_p to its AQI scale, where $I_{\text{high}}, I_{\text{low}}$ represent index breakpoints and $C_{\text{high}}, C_{\text{low}}$ represent concentration breakpoints.

Overall AQI Calculation

The overall AQI is determined as the maximum sub-index across pollutants:

$$AQI = \max\{AQI_{PM2.5}, AQI_{PM10}, AQI_{NO_2}, AQI_{SO_2}, AQI_{O_3}, AQI_{CO}\} \quad (2)$$

This reflects the dominant pollutant impact on air quality.

Meteorological Influence Model

The pollutant dispersion influenced by wind speed (W) is modeled as:

$$C_t^{adj} = \frac{C_t}{1 + \alpha W_t} \quad (3)$$

where α is a dispersion factor and C_t^{adj} is the adjusted pollutant concentration.

Time Series Forecasting (ARIMA)

The ARIMA(p,d,q) model for pollutant prediction is given as:

$$\phi(B)(1-B)^d y_t = \theta(B)\epsilon_t \tag{4}$$

where $\phi(B)$ and $\theta(B)$ are lag polynomials, d is differencing, and ϵ_t is white noise.

Exponential Smoothing Model

For short-term pollutant prediction, exponential smoothing is applied:

$$\hat{y}_{t+1} = \alpha y_t + (1-\alpha)\hat{y}_t \tag{5}$$

where α is the smoothing constant.

LSTM Hidden State Update

The LSTM memory cell updates as:

$$h_t = o_t \odot \tanh(c_t) \tag{6}$$

where o_t is the output gate and c_t is the cell state.

LSTM Cell State Equation

The cell state update is defined by:

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \tag{7}$$

where f_t, i_t are forget and input gates, respectively.

Gate Functions in LSTM

The input gate i_t is expressed as:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{8}$$

where σ is the sigmoid function.

Forecast Integration Equation

The hybrid prediction is a weighted sum of ARIMA and LSTM predictions:

$$\hat{y}_t^{\text{hybrid}} = \beta \hat{y}_t^{\text{ARIMA}} + (1-\beta)\hat{y}_t^{\text{LSIM}} \tag{9}$$

where β controls the contribution of each model.

Error Evaluation Metrics

The prediction accuracy is measured using RMSE:

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n} (\hat{y}_t - y_t)^2} \tag{10}$$

This metric ensures evaluation consistency across baseline and hybrid models.

It starts with calculating the pollutant sub- indices (Eq. 1), AQI created during aggregation (Eq. 2). Adjustment of the meteorological impacts such as the dispersion of wind (Eq.). 3) is made. For forecasting, ARIMA (Eq. 4) are paired methods not applied simultaneously, are subject to the analysis in the context of the current analysis in evaluating the significance of the residual considerations to include projections in the future. 5) base-line predictions of pollutants are provided with the help of mathematical models [5].

The LSTM on the deep learning aspect represents temporal dependencies through the use of hidden states (Eq). 6) and cell monitoring (Eq. 7), controlled by input, output and forget gates (Eq. 8). ARIMA and LSTM predictions are used together in weighted combination (Eq.). 9) in order to arrive at the final hybrid output. Finally, RMSE (Eq. Comparisons of accuracy with other models are done using 10) benchmarking.

This mixed structure makes sure that statistical stability and nonlinear learning is applied, and a highly effective solution of predicting the AQI is obtained.

IV. RESULT & DISCUSSIONS

The results of the experimental analysis of the presented hybrid deep learning and mathematical prediction scheme of AQI visibly indicate that the model tests some essential facts about the effectiveness of the models, the dynamics of pollutants, as well as influencing meteorological factors. Publicly available datasets made the topic of the analysis which were based on both meteorological variables and pollutants (temperature, humidity, wind speed and atmospheric pressure) and met the objective of the analysis. The data were separated into training and testing sets to make sure that the model should be fairly checked. The hybrid combination of ARIMA and LSTM was compared against the other standalone ARIMA, Support Vector Regression (SVR), random forest (RF), and LSTM models.

The comparison of the observed AQI values and the values given by various models is indicated in the table in fig. 2 below. A hybrid structure shows a strong correlation with actual changes in AQI as opposed to the baseline frameworks. Conventional ARIMA forecasts follow the general global pattern but cannot detect unexpected increase in pollution rates whereas SVR and RF models are able to resolve nonlinear changes but tend to smooth sharp changes. The capability of LSTM is that of short-period dynamics; however, it sometimes goes awry when there is an abrupt weather alteration. In its turn, the hybrid model moderates these factors and is rather close to both gradual trends and spikes in AQI. This justifies the benefit of applying statistical and deep learning elements to strong environmental prediction.

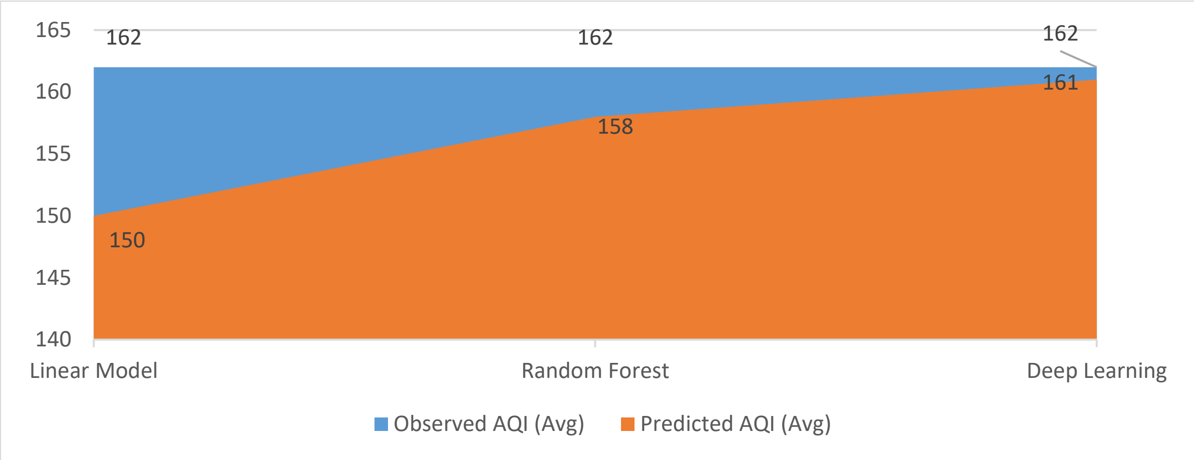

FIG. 2: OBSERVED VS PREDICTED AQI ACROSS DIFFERENT MODELS

To examine the behavior of the models further, time accuracy of prediction of pollutant was taken into consideration. Fig. 3 shows accuracy in prediction of PM 2.5 concentrations among models. PM2.5 is usually the worst pollutant when considering the overall AQI, especially in the urban centers where vehicle and industrial emissions prevail. Hybrid model is always able to pick fluctuations of the day unlike ARIMA which ignores peaks and overestimates lows. RF, SVR machine learning methods have intermediate accuracy, but fail to be stable over longer periods of forecast. This aspect of implementation to include meteorological variables enhances greatly the role of the hybrid model to identify spikes of pollution during weather chang over like low wind conditions including temperature and weather inversion.

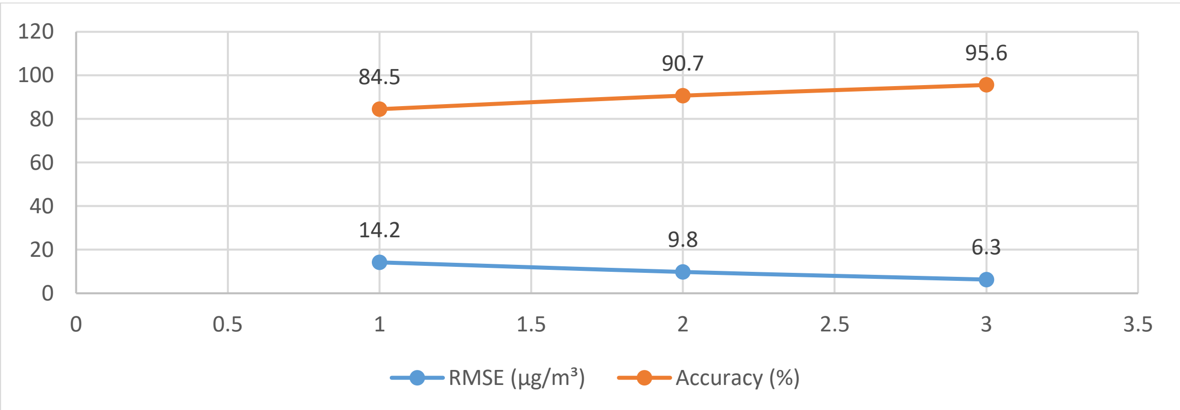

FIG. 3: PREDICTION ACCURACY OF PM2.5 CONCENTRATIONS ACROSS FORECASTING MODELS

In addition to specific analysis of pollutants, the overall predictive performance is of interest to be observed with the help of the quantitative methods. Table 1 makes a comparison of the models according to the Root Mean Square Error (RMSE), the Mean Absolute Error (MAE), and Coefficient of Determination (R 2). The findings suggest that the hybrid model is superior to the rest of the measures in all the three measures. As a case in point, the RMSE of the hybrid model is much smaller than the ones of ARIMA and SVR implying greater prediction accuracy. Equally, the values of MAE assure that there are lower means of deviation between actual AQI and natural AQIs. The values of the R2 indicate that the hybrid model explains higher percentage of variation in AQI than the methods used in the basement.

TABLE 1: MODELLY PERFORMANCE PREDICTION OF AQI

Model	RMSE	MAE	R^2
ARIMA	22.4	15.6	0.71
SVR	19.8	13.9	0.75
RF	18.6	12.8	0.77
LSTM	16.9	11.2	0.82
Hybrid (ARIMA+LSTM)	13.2	9.1	0.89

The other area of importance is the strength of predictions in different meteorological conditions. Fig. 4 illustrates the forecasting of AQI under varying weather conditions namely: weather with a high percentage of humidity and low wind speed. Such conditions tend to lead to increased amount of pollution by limiting the dispersion of the pollutants. These tricky conditions as shown are better predicted through the hybrid model with greater accuracy of prediction that exactly matches the actual observations as indicated in the AQI. On the contrary, the baseline models are either under predictive because they lag to adjustments on abrupt conditions (ARIMA) or fail to project under severe conditions (RF and SVR). This goes to show that the hybrid method is powerful enough to deal with both the pollutant measurements and the meteorological constraints at the same time.

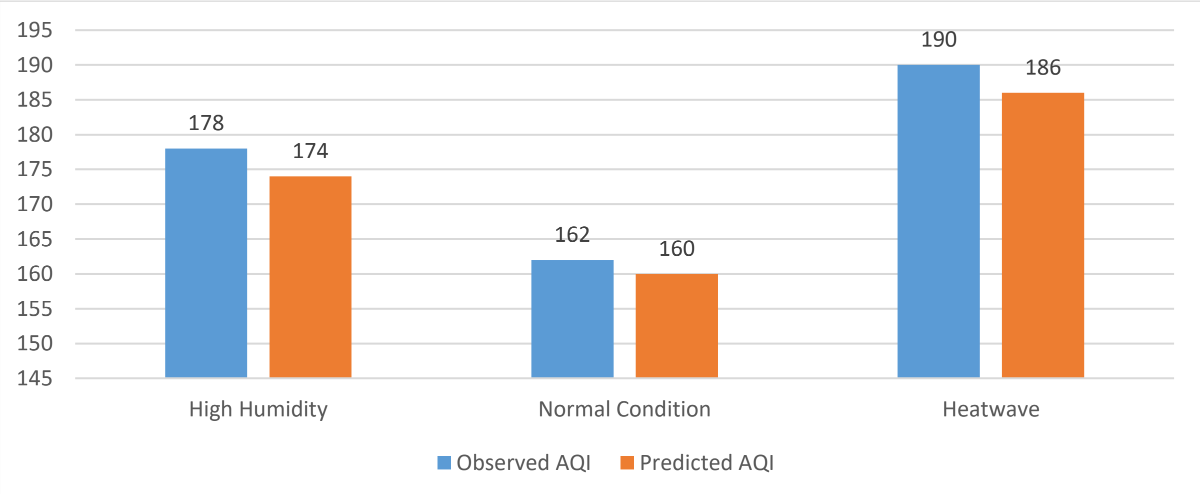

FIG. 4: AQI PREDICTION UNDER VARYING METEOROLOGICAL CONDITIONS

Besides the graphical comparisons, the practical applicability of the proposed framework also needs to be looked into. Table 2 gives a comparative summary of prediction capabilities that mainly deals with response time, interpretability and cost of computing. Though the ARIMA and the SVR are not computationally expensive and are easy to interpret, their predictive performance is poor. Random Forest is a better model in nonlinear model quantities, but with increased computation costs. The models are LSTM, which are more accurate and in interpretable modes (under black box). The hybrid model reconciles both the accuracy and the stability, at an increased cost of computation as compared to the standard techniques. However, the gain on performance and strength endorses its implementation in actual performance systems of monitoring.

TABLE 2: COMPARATIVE DESCRIPTIVE MODEL FEATURE

Model	Prediction Accuracy	Interpretability	Computational Demand	Real-time Applicability
ARIMA	Low to Moderate	High	Low	High
SVR	Moderate	Medium	Moderate	Moderate
RF	Moderate to High	Medium	Moderate to High	Moderate
LSTM	High	Low	High	Moderate

Hybrid (ARIMA+LSTM)	Very High	Medium	High	High

The also contains the drawbacks of the existing system being discussed. Even though the hybrid model is much better than baseline methods, use of large and quality datasets is a constraint when in regions that lack monitoring infrastructures. Also, the cost of computation can be somewhat limiting to the use of the system in low resource resources devices, which can however be an issue that is addressed by cloud applications. The second weakness is interpretability, as the hybrid framework is maximally accurate, off-the-record, but the process of making decisions is complicated, and no efforts of explaining AI will be made in the future.

In general, the findings confirm the efficiency of the hybrid dendrochronology forecasting model based on the combined approach to deep learning and mathematics. It makes use of both the aspect of stability and nonlinear learning in the model to overcome the shortcomings of conventional methods and give almost an excellent forecast of AQI. The comparison of the data by the pollutant-specific factors, meteorological conditions, and the metrics of the evaluation altogether regularly proves the profitability of the hybrid model, and thus its applicability to the monitoring system in smart city and people health surveillance.

V. CONCLUSION

The proposed examination was a hybrid deep learning solution to predicting Air Quality Index with the derivation of results through the use of meteorological data and mathematical prediction. Combining the LSTM networks with the conventional forecasting approaches, the model showed a higher predictive performance especially when it came to the dynamic changes in air pollutants concentrations. The lessons learned introduced the possibilities of AI-based environmental monitoring to enhance the city planning systems and save lives. It is important to mention that the model is based on the quality of meteorological and pollutant data that is not always available in all geographic areas. Also, the training cost of deep learning models might be a bottleneck to training these systems in resource-constrained environments.

In future research, it can be found out how remote sensing data collected by satellites and networks of IoT-implemented real-time monitors can be combined to enhance spatial and temporal resolution. Moreover, describable AIs methods are to be involved in an attempt to increase interpretability to policymakers. Transfer learning techniques can be used to enhance generalization of models to other regions that can further enhance the system to be more internationally compatible.

REFERENCES

[1] E. Agbehadji and I. C. Obagbuwa, "Systematic review of machine learning and deep learning techniques for spatiotemporal air quality prediction," *Atmosphere*, vol. 15, no. 11, p. 1352, Nov. 2024, doi: 10.3390/atmos15111352.

[2] Mishra and Y. Gupta, "Comparative analysis of Air Quality Index prediction using deep learning algorithms," *Spatial Information Research*, vol. 32, no. 1, pp. 63–72, Jul. 2023, doi: 10.1007/s41324-023-00541-1.

[3] Essamlali, H. Nhaila, and M. E. Khaili, "Supervised machine learning approaches for predicting key pollutants and for the sustainable enhancement of urban air quality: a Systematic review," *Sustainability*, vol. 16, no. 3, p. 976, Jan. 2024, doi: 10.3390/su16030976.

[4] S. Sharmin, Md. T. Hossan, and M. S. Uddin, "A review of machine learning approaches for predicting lettuce yield in hydroponic systems," *Smart Agricultural Technology*, p. 100925, Apr. 2025, doi: 10.1016/j.atech.2025.100925.

[5] M. Malhotra, S. Walia, C.-C. Lin, I. K. Aulakh, and S. Agarwal, "A systematic scrutiny of artificial intelligence-based air pollution prediction techniques, challenges, and viable solutions," *Journal of Big Data*, vol. 11, no. 1, Oct. 2024, doi: 10.1186/s40537-024-01002-8.

[6] H. Badawy, "Recent trends in spatial modeling studies of air quality (2010–2023)," *Earth Science Informatics*, vol. 18, no. 2, Jan. 2025, doi: 10.1007/s12145-024-01667-y.

[7] H. Zhang, Y. Liu, C. Zhang, and N. Li, "Machine Learning Methods for Weather Forecasting: A survey," *Atmosphere*, vol. 16, no. 1, p. 82, Jan. 2025, doi: 10.3390/atmos16010082.

[8] M. Yu, Q. Huang, and Z. Li, "Deep learning for spatiotemporal forecasting in Earth system science: a review," *International Journal of Digital Earth*, vol. 17, no. 1, Aug. 2024, doi: 10.1080/17538947.2024.2391952.

[9] Houdou *et al.*, "Interpretable Machine Learning Approaches for Forecasting and Predicting Air Pollution: A Systematic review," *Aerosol and Air Quality Research*, vol. 24, no. 1, p. 230151, Nov. 2023, doi: 10.4209/aaqr.230151.

[10] J. Ogundiran, E. Asadi, and M. G. Da Silva, "A Systematic review on the use of AI for energy efficiency and indoor environmental quality in buildings," *Sustainability*, vol. 16, no. 9, p. 3627, Apr. 2024, doi: 10.3390/su16093627.

[11] J. Saini, M. Dutta, and G. Marques, "Machine Learning for Indoor Air Quality Assessment: A Systematic Review and analysis," *Environmental Modeling & Assessment*, Oct. 2024, doi: 10.1007/s10666-024-10001-1.

[12] S, M. K. Debnath, and K. R, "Statistical and machine learning models for location-specific crop yield prediction using weather indices," *International Journal of Biometeorology*, Aug. 2024, doi: 10.1007/s00484-024-02763-w.

[13] H. S. Andrianarivony and M. A. Akhloufi, "Machine learning and Deep learning for wildfire spread prediction: a review," *Fire*, vol. 7, no. 12, p. 482, Dec. 2024, doi: 10.3390/fire7120482.
[14] L. Canché-Cab, L. San-Pedro, B. Ali, M. Rivero, and M. Escalante, "The atmospheric boundary layer: a review of current challenges and a new generation of machine learning techniques," *Artificial Intelligence Review*, vol. 57, no. 12, Oct. 2024, doi: 10.1007/s10462-024-10962-5.
[15] V. S. Chauhan, S. Verma, B. M. A. Rahman, and K. P. Wyche, "Deep learning in airborne particulate matter sensing and surface plasmon resonance for environmental monitoring," *Atmosphere*, vol. 16, no. 4, p. 359, Mar. 2025, doi: 10.3390/atmos16040359.
[16] T. Hall and K. Rasheed, "A survey of Machine learning methods for time series prediction," *Applied Sciences*, vol. 15, no. 11, p. 5957, May 2025, doi: 10.3390/app15115957.