International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com
© 2026 IJERST. All Rights Reserved
Peer Reviewed, Referred & Indexed Journal

Research Paper

# EMPLOYING GENETIC ALGORITHMS FOR PREDICTING AIR QUALITY INDEXES THROUGH IMPROVED EXTREME LEARNING MACHINES

Md. Ateeq Ur Rahman
Professor, Department of Computer Science and Engineering,
Shadan College of Engineering and Technology,
Hyderabad, Telangana, India – 500086.
Email: mail_to_ateeq@yahoo.com

Subramanian K.M
Professor, Department of Computer Science and Engineering,
Shadan College of Engineering and Technology,
Hyderabad, Telangana, India – 500086.
Email: kmsubbu.phd@gmail.com

***Abstract*— Properly predicting air quality is crucial to protecting public health and shaping environmental policies. The traditional single model approaches are typically poor in predicting the variations of air quality. In order to solve this, the present study suggests a powerful prediction method based on advanced machine learning methods. We analyse comparatively several models such as SVR, GA-KELM, DBN-BP. Moreover, we propose BiLSTM, a DL model to improve the prediction accuracy. We validate the performance of BiLSTM on a large number of experiments and analyses, proving the superiority of BiLSTM over the existing state-of-the-art models in terms of reduced RMSE value, reduced MSE value. In addition, we optimize the performance of the BiLSTM by GA-KELM, further improving the predictive power of the system. The suggested hybrid model not only ensures better accuracy in forecasting air quality but also helps in decision making for pollution control and public health measures. This research emphasises the need to explore new ways for solving important environmental problems and the potential role of ML in improving air quality management.**

***"Keywords— Time series, air quality forecasting, machine learning, extreme learning machine, genetic algorithm."***

## I. INTRODUCTION

In 21st century, pollution in the air becomes an important problem in the whole world, due to industrialisation and urbanisation [1]. The decreasing air quality has far-reaching impacts on the environment and public health [2]. The health impacts of outdoor physical activity on exposure to ambient air pollution, especially in areas of rapid industrialization (like China) are underlined in the work by Li et al. [3]. The air quality in China (and in many other countries) is evaluated by using the air quality parameters provided in the Chinese Ambient Air Quality Standards including SO2, NO2, PM10 and 2PM2.5 . O3, and CO [4].

Health hazards of these contaminants are widely recognised [5]. The International Energy Agency estimates that air pollution causes an estimated 6.5 million premature deaths annually and long-term exposure to air pollution, including air pollutants such as PM2.5 and traffic-related emissions, is associated with increased incidence of lung cancer, coronary heart disease, and other diseases [6]. Therefore, it is necessary to create efficient air quality forecast systems to be an integral part of the work on protection of the environment [7].

The forecasting of air quality is primarily based on information from the monitoring stations of the major cities [8]. These stations are very useful for obtaining pollution data and are also helpful to feed into algorithms for forecasting. ML algorithms have been found to be useful tools to leverage such data for analysis and to automatically learn features at different levels of abstraction [9]. There are still some challenges however, including the lack of comprehensive and detailed information, and the difficulty of modelling many contaminants simultaneously [10].

There are many techniques that have been investigated recently to address these challenges. The data from six atmospheric contaminants were used by Wu et al. to suggest an optimal hybrid model for forecasting the daily AQI [11]. Typically, the classic neural network algorithms face issues like slow learning, local minimal and complicated training procedure [12].

Huang et al. proposed the ELM technique based on the generalised inverse matrix theory and has a single hidden layer feedforward neural network to solve these restrictions [13]. Compared to conventional NN, the ELM approach is shown to have superior performance for AQI prediction, better parameter selection, training time and prediction accuracy [14]. But the ELM algorithm has a disadvantage of relying on the randomness of the hidden layer nodes, thus affecting the accuracy of prediction [15].

There are some limitations to the current air quality prediction models. The proposed strategy in this research is a revolutionary approach that merges the power of the ML algorithms and enhanced parameters optimization techniques. In particular, we design a hybrid model, which is a combination of BiLSTM and GA-KELM. This is a combination to enhance the predictive accuracy and robustness of air quality forecasts using the predictive

potential of BiLSTM and optimizing the parameters of the model using the genetic algorithm [16].

Finally, the current study is an attempt to fill the gaps of previous studies on the improvement of Air Quality Forecasting methods by proposing a novel hybrid method. This paper aims to develop an accurate and reliable forecasting model that can be used as a basis for environmental protection and public health management decision making, which is achieved by adopting BiLSTM and GA-KELM.

## II. Related Work

It is an important environmental and public health problem and there is a need for substantial research to gain knowledge about the sources, effects and reduction methods for air pollution on an international level. This literature study aims at presenting the major literature relevant to air pollution monitoring, forecasting and control, especially with regard to the employment of ML methods for air quality forecasting.

It is crucial to address the air pollution problem at regional level as regional air pollution is a worldwide issue that require regional solutions, as stated by Li et al. (2019) [1]. Such a perspective highlights the need for localised air quality monitoring and forecast systems to guide specific measures. Han et al. Advanced statistical methods have been used for air quality data analysis, such as a Bayesian LSTM model used to evaluate the impact of air pollution control measures in China in 2018. They focus on how LSTM models can predict the impacts of policy actions on AQ outcomes.

Bai et al. (2018) reviewed the literature on air pollution forecasts, covering the different approaches used to model and the data sources used for prediction of air quality [3]. Their evaluation shows that air quality forecasts are complex and require many data inputs, such as meteorological data, satellite observations and ground-based monitoring data. Xue and Ding (2019) came up with a DL approach to writer identification using air handwriting data from inertial sensors and showcased the versatility of DL techniques for processing sensor data for other applications [4].

Cheng et al. (2019) researched the optimization of the outside air ratio in air conditioning systems aiming to achieve a specific indoor air quality and maximum energy saving [5]. Their findings further underscores the need to optimise ventilation strategies for indoor air quality and energy savings. Air pollutants prediction in major cities of India using time series based LSTM model has been proposed by Chaudhary et al [6] demonstrating the usefulness of LSTM model in air quality prediction. Their work builds on a growing body of work focused on data-centric strategies and methods for predicting air quality.

Chen et al. (2018) outlined a framework for urban healthcare big data system that leverages crowdsourced and cloud-based air quality metrics, highlighting the effectiveness of crowdsourced data in urban air quality monitoring [7]. Their efforts have shown the potential of new technology to improve air quality monitoring and public health surveillance. Du and colleagues. (2021) proposed a deep air quality forecasting framework with a hybrid DL method, which integrated the CNNs and LSTM networks [8]. Their study reveals that hybrid DL models are excellent in collecting complicated spatiotemporal patterns in air quality data.

Overall, the literature analysis shows the increasing interest in the use of ML approaches for air quality monitoring, forecasting and management. There have been studies on a wide variety of ways to enhance the accuracy and reliability of air quality predictions including LSTM models , DL frameworks , and hybrid ML architectures . Such advances are promising for influencing evidence-based actions to minimise the detrimental consequences of air pollution on public health and the environment.

## III. Materials And Methods

The proposed plan is designed to combine ELM and GA to enhance the air quality prediction especially predicting PM2.5. The selection of the hidden nodes and layers of the ELM model will be optimised using GA so as to improve the learning ability and prediction accuracy. Therefore, ELM model can alter the architecture to fit the complicated relationship in air quality data more effectively by using GA[14] to search for the best solution in a given search space. The comparative analysis will be carried out with the existing methods like SVM [16] etc. MSE and RMSE will be used to assess the effectiveness of the performance metrics. The proposed solution proposes to provide a more reliable and accurate air quality forecasting system that would allow for a wide-scale assessment of air quality and potential impacts on air quality on public health and the environment.

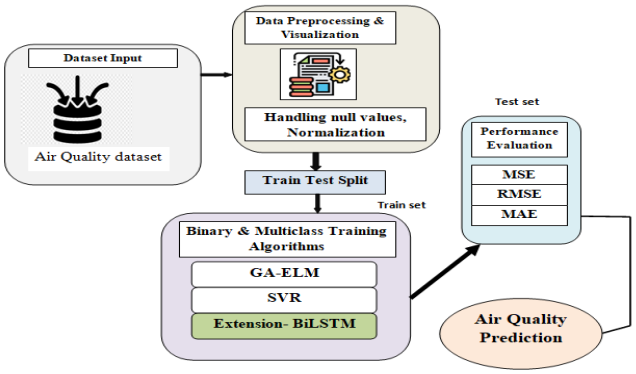



Fig.1 Proposed Architecture

The architecture of air quality prediction system has numerous important elements to efficiently process, train and evaluate predictive models.

*Dataset Input:* Air quality data is fed into the system, which includes relevant data like pollutant concentrations, weather conditions and location information.

*Data Processing and Visualization:* In this step, we do pre-processing such null value management, normalisation and feature engineering to prepare the data for modelling. Data distributions and correlations are learned using visualisation techniques.

*Train-Test Split:* Dividing the data into a training dataset and a testing dataset for training and evaluation. This way, you can test the model's performance on new data to assess its generalisation ability.

*Binary and Multi-Class Training Algorithms:* The commonly used techniques for binary and multi-class training are GA-ELM [14], SVR [16] and BiLSTM networks. These algorithms are used to teach the training data the basic patterns and relationships between the input variables and the air quality outcomes.

*Performance Evaluation:* The effectiveness of the model is evaluated by MSE, RMSE and MAE. These metrics measure the difference between the expected and actual air quality values and they provide a measure of the accuracy and reliability of air quality models.

*Air Quality Prediction:* The created models can then be used to predict on new data, predicting air quality parameters including pollutant exposure or AQI scores. These forecasts are essential to understanding the current and forecasted air quality situation, and informing decisions on how to manage pollution and public health.

To sum up, the system design offers a thorough approach for air quality forecasting, utilizing machine learning algorithms and performance evaluation techniques to deliver accurate and reliable predictions.

*A) Dataset Collection:*

The air quality dataset consists of parameters like sulphur dioxide (SO2), nitrogen dioxide (NO2), particulate matter less than 10 microns (PM10), less than 2.5 microns (PM2.5), ozone (O3) and carbon monoxide (CO). Each observation contains the concentrations of pollutants, the timing of the measurements, and the geographic coordinates of the measurements. Additional data like temperature, humidity, wind velocity and atmospheric pressure can also be incorporated. The dataset can be used to look at and understand air quality trends over time, and from region to region, as well as to analyze the effects of air pollution on the environment and human health.

| | City | Date | PM2.5 | PM10 | NO | NO2 | NOx | NH3 | CO | SO2 | O3 | Benzene | Toluene | Xylene | AQI | AQI_Bucket |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Ahmedabad | 2015-01-01 | 0.0 | 0.0 | 0.92 | 18.22 | 17.15 | 0.0 | 0.92 | 27.64 | 133.36 | 0.00 | 0.02 | 0.00 | 0.0 | 0 |
| 1 | Ahmedabad | 2015-01-02 | 0.0 | 0.0 | 0.97 | 15.69 | 16.46 | 0.0 | 0.97 | 24.55 | 34.06 | 3.68 | 5.50 | 3.77 | 0.0 | 0 |
| 2 | Ahmedabad | 2015-01-03 | 0.0 | 0.0 | 17.40 | 19.30 | 29.70 | 0.0 | 17.40 | 29.07 | 30.70 | 6.80 | 16.40 | 2.25 | 0.0 | 0 |
| 3 | Ahmedabad | 2015-01-04 | 0.0 | 0.0 | 1.70 | 18.48 | 17.97 | 0.0 | 1.70 | 18.59 | 36.08 | 4.43 | 10.14 | 1.00 | 0.0 | 0 |
| 4 | Ahmedabad | 2015-01-05 | 0.0 | 0.0 | 22.10 | 21.42 | 37.76 | 0.0 | 22.10 | 39.33 | 39.31 | 7.01 | 18.89 | 2.78 | 0.0 | 0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 497 | Ahmedabad | 2016-05-12 | 0.0 | 0.0 | 0.00 | 0.00 | 0.00 | 0.0 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0 | 0 |
| 498 | Ahmedabad | 2016-05-13 | 0.0 | 0.0 | 0.00 | 0.00 | 0.00 | 0.0 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0 | 0 |
| 499 | Ahmedabad | 2016-05-14 | 0.0 | 0.0 | 0.00 | 0.00 | 0.00 | 0.0 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0 | 0 |
| 500 | Ahmedabad | 2016-05-15 | 0.0 | 0.0 | 0.00 | 0.00 | 0.00 | 0.0 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0 | 0 |
| 501 | Ahmedabad | 2016-05-16 | 0.0 | 0.0 | 0.00 | 0.00 | 0.00 | 0.0 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.0 | 0 |

"Fig.2 Dataset Collection"

*B) Processing:*

*Data Processing with Pandas DataFrame:* Data Transformation activities and activity preparation will be performed on Pandas DataFrame. This involves filling in missing data, normalising data, and removing unnecessary columns from the data to make it suitable for model training.

*Handling Missing Values:* Methods for Imputation and Exclusion is used for dealing with missing data. This guarantees the dataset's comprehensiveness and prevents future research from being affected by this.

*Normalization:* By adjusting numerical attributes to a uniform scale usually between 0 and 1, characteristic scales within the data set are removed from the modelling process where they may have very different ranges.

*Dropping Unwanted Columns:* Columns that are not useful for the predictive analysis are removed from the DataFrame. This helps to lower dimension and improve computation efficiency when training the model.

*Data Processing with Keras DataFrame:* Keras DataFrame provides a user-friendly way to interact with deep learning frameworks, enabling fast data preprocessing and training of deep learning models for neural network architectures.

*Handling Missing Values:* Missing values are handled the same as in Pandas to ensure a complete and coherent data set.

*Normalization:* Numerical attributes of the Keras DataFrame are scaled properly using normalisation techniques. This helps the neural network to converge during training by keeping the attributes in the same range.

*Dropping Unwanted Columns:* The Keras DataFrame class will mark those columns that it considers to be unnecessary for training a neural network. This reduces model complexity, hence optimising computational resources and averting overfitting.

Data processing is a critical step in data preparation for model training, where both Pandas and Keras DataFrame is used for data processing, improving data quality and enabling fast model convergence.

*Visualization:* The Seaborn and Matplotlib packages can be used to visualize the air quality data. This will help improve our understanding of the material, as it will be presented in a visual format.

*Histograms:* A histogram from the Seaborn libary's histplot function and Matplotlib's hist function is used to detect any patterns or irregularities in the pollutant levels.

*Scatter Plots:* Seaborn and Matplotlib provide scatterplot functionality to visualize scatterplots for different pollutants or correlation between pollutants and meteorological data, which helps to determine correlation.

*Line Plots:* For visualizing the changes in pollutant concentration with time, use the Seaborn lineplot and Matplotlib plot functions to create Line Graphs. These charts can be helpful for trend analysis and seasonal analysis.

These visualizations provide useful information on the temporal changes in air quality, which will be useful for ongoing research and development of air quality models.

*Feature Selection:* It is crucial to select features to build successful air quality prediction model. Technique used include correlation analysis, feature importance ranking, dimensionality reduction technique like PCA. Correlation analysis can be used to help determine the correlation between each contaminant and climatic variable which will help to select the relevant ones. Ranking methods e.g., RF feature importance, rank features by the size of the effect they have on the prediction. PCA is also used to discover 'latent variables' that explain the bulk of the variation within the data whilst preserving key information, thereby reducing the number of variables. Feature selection enhances the efficiency and effectiveness of the models by selecting the most relevant features for air quality prediction problems.

*Training & Testing:* It is important to partition the air quality dataset into training and testing sets to determine the effectiveness of the model. In most cases this random split occurs (e.g. 80/20 or 70/30) meaning there is enough data to train and test the machine learning model. The training data set is used to build the prediction model; the test data set is not used in building the prediction model, but only to check the effectiveness of the prediction model trained. This split enables us to assess how well a model performs on previously

unseen data, and prevent overfitting to the evaluation data set, leading to greater confidence in the model's performance in actual applications.

*C) Algorithms:*

*Genetic Algorithm with Extreme Learning Machine (GA-ELM):* GA-ELM is the fusion of evolutionary optimisation methodology of GA [14] and the efficient learning methodology of ELM. In GA-ELM, GA is used to optimise the parameters of the ELM model so as to improve the prediction performance. The GA starts with an initial population of potential ELM solutions, which are selected, crossed, and mutated until a "validation set" is used to evaluate their fitness, and the most successful individuals are passed on. In contrast, for the ELM, the input features are mapped to a higher dimensional space with one hidden layer of neurons with randomly generated non-linear activations. Then the output weights are computed using the Moore-Penrose pseudoinverse.

*Support Vector Regressor (SVR):* SVR fits a regression model based on the notion of structural risk minimisation. It attempts to find the hyperplane which is best separating the data points with the widest margin. During model training, SVR[16] searches for the parameters of the model (such as kernel, regularisation parameter) that minimise the loss function, which is usually an epsilon-insensitive loss function. Once trained, SVR models can be used to predict air quality ratings for unseen data, thanks to its ability to capture complex air quality patterns.

The equation of the linear hyperplane can be expressed as:

wTx+b=0$wTx+b$=0 (1)

Where:

- The vector ww is normal to the hyperplane (the direction perpendicular to the hyperplane).
- The offset or bias term is denoted as bb. The distance of the hyperplane from the origin in the direction of the normal vector ww.

*Bidirectional Long Short-Term Memory (BiLSTM):* BiLSTM is an extension of LSTM neural network architecture which can capture long-range relationships in sequential input. The BiLSTM is able to learn from both past and future context, since it takes in input sequences in both directions. This ability is quite useful for modelling the temporal trend in air quality data as past and future observations can affect current air quality. BiLSTM models are ideal for air quality prediction tasks since they have been shown to be effective in capturing complicated temporal dynamics.

## IV. Experimental Results

MSE: MSE is one of the measures of error in statistical models. It is the mean of the squares of the deviations between the observed values and the projected values. If the model has no mistake then MSE is 0. The greater the inaccuracy of the model, the higher its value. Also called MSD, MSE is.

$$\text{"MSE} = \frac{\Sigma(\text{yi} - \text{pi})2}{\text{n}} \text{(2)"}$$

RMSE: RMSE is the typical difference between the values of a statistical model and the real values. Mathematically, it is the standard deviation of the residuals. The distances of the points from the regression line are called residuals.

$$\text{“RMSE} = \text{sqrt } [(\Sigma(\text{Pi} - \text{Oi})^2) / \text{n}] \quad \text{(3)”}$$

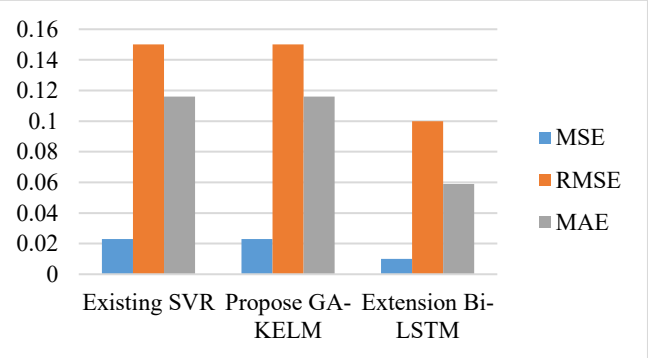
MAE: Absolute mistake is how much mistake is in your measurements. The difference between the measured value and the "true" value. For instance, if the scale reads 90 pounds and you know that you weigh 89 pounds, then the absolute inaccuracy of the scale is 90 lbs - 89 lbs = 1 lbs.

$$\text{MAE} = (1/n) * \Sigma \text{ |Actual - Predicted|} \quad (4)$$

The performance measures, namely, Accuracy, precision, Recall, and F1-score of each algorithm are presented in table (1). The BILSTM extension always performs better than all other algorithms. Tables also display comparisons of the metrics for each of the algorithms

Table.1 Performance Evaluation Table

| ML Model | MSE | RMSE | MAE |
|---|---|---|---|
| Existing SVR | 0.023 | 0.15 | 0.116 |
| Propose GA-KELM | 0.023 | 0.15 | 0.116 |
| Extension Bi-LSTM | 0.010 | 0.10 | 0.059 |




Graph.1 Comparison Graphs

Graph (1): MSE (blue), RMSE (orange), MAE (grey). Extension BILSTM model shows the best performance in all the measures, which scores are higher than the other models. The results are shown in a graph above.

## V. Conclusion

In summary, GA-KELM and its extension of BiLSTM are significant developments towards air quality prediction, offering enhanced prediction accuracy and advancing the process of decision making for environmental management. The implementation of BiLSTM model in a user-friendly Flask framework further adds to the effect of the project, giving real-time access to air quality predictions to researchers and general public. This would enable people to make educated decisions regarding their health and well-being and to take proactive steps to limit the bad impacts of air pollution on the environment.

Several future research and development directions are apparent. Firstly, the GA-KELM and BiLSTM models can be further optimised and improved to increase the accuracy of predictions and resilience. In addition, real-time data streams can be added and a larger number of attributes are possible. In addition, other uses of the models, beyond air quality prediction, like climate modelling or environmental impact assessment, would be of interest. Finally, more accessible and user-friendly technologies for predicting air quality, such as mobile applications and web applications, can help expand the use and usefulness of these technologies to improve environmental problems. To summarise, the future scope is to enhance computational approaches and use them to handle the changing environmental issues and improve the quality of life around the globe.

## REFERENCES


[1] X. Li, L. Jin, and H. Kan, ''Air pollution: A global problem needs local fixes,'' Nature, vol. 570, no. 7762, pp. 437–439, Jun. 2019.

[2] Y. Han, J. C. K. Lam, and V. O. K. Li, ''A Bayesian LSTM model to evaluate the effects of air pollution control regulations in China,'' in Proc. IEEE Big Data Workshop (Big Data), Dec. 2018, pp. 4465–4468.

[3] L. Bai, J. Wang, X. Ma, and H. Lu, ''Air pollution forecasts: An overview,'' Int. J. Environ. Res. Public Health, vol. 15, no. 4, p. 780, 2018.

[4] Y. Ding and Y. Xue, ''A deep learning approach to writer identification using inertial sensor data of air-handwriting,'' IEICE Trans. Inf. Syst., vol. E102-D, no. 10, pp. 2059–2063, 2019.

[5] S.-Q. Dotse, M. I. Petra, L. Dagar, and L. C. De Silva, ''Application of computational intelligence techniques to forecast daily PM10 exceedances in Brunei Darussalam,'' Atmos. Pollut. Res., vol. 9, no. 2, pp. 358–368, Mar. 2018.

[6] M. Jia, A. Komeily, Y. Wang, and R. S. Srinivasan, ''Adopting Internet of Things for the development of smart buildings: A review of enabling technologies and applications,'' Automat. Construct., vol. 101, pp. 111–126, May 2019.

[7] S. Abirami, P. Chitra, R. Madhumitha, and S. R. Kesavan, ''Hybrid spatio-temporal deep learning framework for particulate matter (PM2.5) concentration forecasting,'' in Proc. Int. Conf. Innov. Trends Inf. Technol. (ICITIIT), Feb. 2020, pp. 1–6.

[8] Y. Cheng, S. Zhang, C. Huan, M. O. Oladokun, and Z. Lin, ''Optimization on fresh outdoor air ratio of air conditioning system with stratum ventilation for both targeted indoor air quality and maximal energy saving,'' Building Environ., vol. 147, pp. 11–22, Jan. 2019.

[9] A. C. Cosma and R. Simha, ''Machine learning method for real-time non-invasive prediction of individual thermal preference in transient conditions,'' Building Environ., vol. 148, pp. 372–383, Jan. 2019.

[10] M. Bhowmik, K. Deb, A. Debnath, and B. Saha, ''Mixed phase Fe2O3/Mn3O4 magnetic nanocomposite for enhanced adsorption of methyl orange dye: Neural network modeling and response surface methodology optimization,'' Appl. Organometallic Chem., vol. 32, no. 3, p. e4186, Mar. 2018.

[11] V. Chaudhary, A. Deshbhratar, V. Kumar, and D. Paul, ''Time series based LSTM model to predict air pollutant's concentration for prominent cities in India,'' in Proc. Int. Workshop Utility-Driven Mining (UDM), Aug. 2018, pp. 1–9.

[12] M. Chen, J. Yang, L. Hu, M. S. Hossain, and G. Muhammad, ''Urban healthcare big data system based on crowdsourced and cloud-based air quality indicators,'' IEEE Commun. Mag., vol. 56, no. 11, pp. 14–20, Nov. 2018.

[13] R. Chen, X. Wang, W. Zhang, X. Zhu, A. Li, and C. Yang, ''A hybrid CNN-LSTM model for typhoon formation forecasting,'' GeoInformatica, vol. 23, no. 3, pp. 375–396, Jul. 2019.

[14] S. Du, T. Li, Y. Yang, and S. Horng, ''Deep air quality forecasting using hybrid deep learning framework,'' IEEE Trans. Knowl. Data Eng., vol. 33, no. 6, pp. 2412–2424, Jun. 2021.

[15] R. Feng, H.-J. Zheng, H. Gao, A.-R. Zhang, C. Huang, J.-X. Zhang, K. Luo, and J.-R. Fan, ''Recurrent neural network and random forest for analysis and accurate forecast of atmospheric pollutants: A case study in Hangzhou, China,'' J. Cleaner Prod., vol. 231, pp. 1005–1015, Sep. 2019.

[16] B. S. Freeman, G. Taylor, B. Gharabaghi, and J. Thé, ''Forecasting air quality time series using deep learning,'' J. Air Waste Manage. Assoc., vol. 68, no. 8, pp. 866–886, Aug. 2018.

[17] S. Mahajan, H.-M. Liu, T.-C. Tsai, and L.-J. Chen, ''Improving the accuracy and efficiency of PM2.5 forecast service using cluster-based hybrid neural network model,'' IEEE Access, vol. 6, pp. 19193–19204, 2018.

[18] J. Jin, J. Gubbi, S. Marusic, and M. Palaniswami, ''An information framework for creating a smart city through Internet of Things,'' IEEE Internet Things J., vol. 1, no. 2, pp. 112–121, Apr. 2014.

[19] A. Grover, A. Kapoor, and E. Horvitz, ''A deep hybrid model for weather forecasting,'' in Proc. 21st ACM SIGKDD Int. Conf. Knowl. DiscoveryData Mining, Aug. 2015, p. 379–386.

[20] A. Agarwal and M. Sahu, ''Forecasting PM2.5 concentrations using statistical modeling for Bengaluru and Delhi regions,'' Environ. Monit. Assessment, vol. 195, p. 502, Mar. 2023.

[21] D. J. Lary, T. Lary, and B. Sattler, ''Using machine learning to estimate global PM2.5 for environmental health studies,'' Environ. Health Insights, vol. 9, no. 1, pp. 41–52, 2015.

[22] Y. Zheng, X. Yi, M. Li, R. Li, Z. Shan, E. Chang, and T. Li, ''Forecasting fine-grained air quality based on big data,'' in Proc. 21th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, New York, NY, USA, Aug. 2015, pp. 2267–2276.

[23] Y. Zheng, L. Capra, O. Wolfson, and H. Yang, ''Urban computing: Concepts, methodologies, and applications,'' ACM Trans. Intell. Syst. Technol., vol. 5, no. 3, p. 38, Sep. 2014.

[24] T. S. Rajput and N. Sharma, ''Multivariate regression analysis of air quality index for Hyderabad city: Forecasting model with hourly frequency,'' Int. J. Appl. Res., vol. 3, no. 8, pp. 443–447, 2017.

[25] Z. Kang and Z. Qu, ''Application of BP neural network optimized by genetic simulated annealing algorithm to prediction of air quality index in Lanzhou,'' in Proc. IEEE Comput. Intell. Appl. (ICCIA), Sep. 2017, pp. 155–160, doi: 10.1109/CIAPP.2017.8167199.

[26] B. Liu, S. Yan, J. Li, G. Qu, Y. Li, J. Lang, and R. Gu, ''A sequence-tosequence air quality predictor based on the n-step recurrent prediction,'' IEEE Access, vol. 7, pp. 43331–43345, 2019.

[27] K. Elbaz, I. Hoteit, W. M. Shaban, and S.-L. Shen, ''Spatiotemporal air quality forecasting and health risk assessment over smart city of NEOM,'' Chemosphere, vol. 313, Feb. 2022, Art. no. 137636.

[28] P. C. Campbell, Y. Tang, P. Lee, B. Baker, D. Tong, R. Saylor, A. Stein, J. Huang, H.-C. Huang, E. Strobach, J. McQueen, L. Pan, I. Stajner, J. Sims, J. Tirado-Delgado, Y. Jung, F. Yang, T. L. Spero, and R. C. Gilliam, ''Development and evaluation of an advanced national air quality forecasting capability using the NOAA Global Forecast System version 16,'' Geosci. Model Develop., vol. 15, no. 8, pp. 3281–3313, Apr. 2022.

[29] X.-B. Jin, Z.-Y. Wang, W.-T. Gong, J.-L. Kong, Y.-T. Bai, T.-L. Su, H.-J. Ma, and P. Chakrabarti, ''Variational Bayesian network with information interpretability filtering for air quality forecasting,'' Mathematics, vol. 11, no. 4, p. 837, Feb. 2023.

[30] J. Li, W. Shi, and D. Yang, ''Color difference classification of dyed fabrics via a kernel extreme learning machine based on an improved grasshopper optimization algorithm,'' Color Res. Appl., vol. 46, no. 2, pp. 388–401, Apr. 2021.