Air Quality Monitoring Using Machine Learning Techniques

Er.K.P.Sridevi
Dept. of Computer Science
VPMM College of Engineering and Technology,
Krishnankovil,Virudhunagar.,India
sridevihodct@gmail.com

Mr.R.Rajkumar.,M.E,MISTE
Associate Professor: dept. of Computer Science
VPMM College of Engineering and Technology,
Krishnankovil,Virudhunagar.,India
hr.rajkumar11@gmail.com

Abstract-**Air pollution (AP) is a major global environmental issue, drawing attention from researchers due to its impact on human health. Predicting air quality is crucial to inform people about health risks and protect against the effects of air pollution, particularly in metropolitan cities facing severe environmental challenges. Real-time monitoring of pollution data enables local authorities to analyze traffic conditions and make informed decisions. In this study, we developed a machine learning model to forecast the air quality index of India, which is a standard measure of pollutant levels (e.g., SO2, NO2) over a specific period. Our model is based on historical data from previous years and uses Gradient Descent-boosted multivariable regression to predict the air quality index for an upcoming year. To improve model efficiency, we applied cost estimation to the predictive problem. The proposed model demonstrates high accuracy, achieving 96% on the current available dataset for predicting India's air quality index. Moreover, we utilized XG Boost and Light GBM algorithms to determine the order of preference based on similarity to the ideal solution, further enhancing the model's performance. Our model has the capability to predict the air quality index for an entire county, state, or any bounded region, given historical pollutant concentration data. By implementing parameter-reducing formulations, we outperformed standard regression models, making our approach a valuable tool for air quality prediction and environmental decision-making to protect human health.**

Keywords—Machine Learning, Air Quality, Air quality prediction, Monitoring System, Intelligent System.

I. Introduction

In recent years, there has been a growing awareness of the significant impact of environmental factors on human health. Air quality has emerged as a central concern in people's daily lives, and there is a strong desire for real-time information on the quality of the air we breathe [1]. However, merely providing real-time data is not enough; it is equally crucial to be able to predict future trends in air pollutants. Currently, weather forecast data has attained a high level of reliability and accuracy, making it an invaluable resource. Building on this, we propose a novel approach that combines predictive weather data with existing historical air quality and meteorological data. To achieve this, we will employ machine learning techniques, which will help us explore data correlations and construct a well-performing model to predict future air quality conditions. By integrating these various data sources, we can develop an efficient solution for air quality prediction, with a focus on feature exploration through predictive data.

The primary goal of air quality prediction is to estimate the concentration of pollutants for a specific time in the future. To achieve this, we rely on historical air quality datasets, meteorological datasets, and other relevant information, as seen in previous research works [2] and [3]. However, we have identified that existing methods predominantly rely on historical data-based predictions using different machine learning approaches, such as neural networks like LSTM [4], machine learning-based solutions [5] [6], Extreme Learning Machine (ELM) [7], or simple regression methods. Our proposed approach aims to build upon the existing methods by harnessing the power of machine learning and leveraging the accurate weather forecast data. By doing so, we can gain deeper insights into data correlations and better understand how various factors influence air quality. This will enable us to construct a more robust and efficient model for predicting air pollution levels accurately.The implications of this research are profound, as improved air quality prediction can have significant positive impacts on public health and environmental protection. By forecasting future air quality conditions, policymakers and the public can proactively take measures to reduce exposure to harmful pollutants, leading to healthier communities and a more sustainable environment.

The current methods for air quality prediction have certain limitations that need to be addressed. Firstly, they do not fully leverage the vast amount of available air quality big data to deeply explore temporal and statistical features. This means that valuable patterns and insights present in the data may remain untapped, limiting the potential for accurate predictions. Secondly, using simple regression methods is not efficient when dealing with high-dimensional big data. These methods may struggle to handle the complexity of the data, leading to lower accuracy in the model's predictions.

Given these shortcomings, further research is essential to develop more sophisticated and advanced techniques for air quality prediction. One promising aspect is the increasing accuracy of meteorological data measurements and the growing reliability of predictive data. These developments offer significant potential for valuable data mining opportunities. Effectively integrating predictive data with historical data holds the key to improving the prediction accuracy. By combining the insights from predictive data with the knowledge from historical records, we can create a more comprehensive and powerful prediction model. In this context, employing a photogrammetric-based method becomes crucial for parameter evaluation, although it comes with the challenge of handling high volumes of data. However, this approach can potentially lead to more accurate parameter evaluations and significantly enhance the overall prediction performance.

To sum up, this paper highlights the necessity to address the limitations of current air quality prediction methods. Utilizing the abundance of air quality big data and effectively combining predictive and historical data is crucial for developing more advanced models. With improved meteorological data and the reliability of predictive data, there is considerable potential for valuable mining insights. Emphasizing the integration of predictive data and employing advanced evaluation methods will pave the way for more accurate and effective air quality predictions, contributing to better environmental and public health management.

Datasets from different sources would be combined to form a generalized dataset, and then different machine learning algorithms would be applied to extract patterns and to obtain results with maximum accuracy. Due to human activities, industrialization and urbanization air is getting polluted. The major air pollutants are CO, NO, C6H6, etc. The concentration of air pollutants in ambient air is governed by the meteorological parameters such as atmospheric wind speed, wind direction, relative humidity, and temperature. Earlier techniques such as Probability, Statistics etc. were used to predict the quality of air.

As population increases so does there is increase in resources used by the human beings. As more the resources are used so there is massive increase of pollution. This main pollution which is most harmful to human body is air pollution. If the quality of the air is very bad the humans can catch diseases. So, the air quality index has become a discussing point in many forums and many researches has tried to suggest a way to alert humans about hazardous air quality index. So there arises a need to address air pollution menace and try to alert a user of bad air quality index using machine learning algorithms as they are powerful enough to predict air quality index by studying historical data of a place.

II. Literature Survey

At present, the methods for air quality prediction are mainly based on simple regression models or neural networks [16]. For example, Zheng et al. [17] introduced a hybrid prediction method that combines two distinct approaches: a linear regression-based temporal prediction method and an artificial neural network (ANN)-based spatial prediction method. This innovative combination aims to capitalize on the strengths of each technique. The linear regression-based temporal prediction method is likely utilized to identify and capture temporal trends and patterns in the air quality data. On the other hand, the ANN-based spatial prediction method is designed to consider spatial dependencies and variations in pollutant concentrations. By integrating these two methods, the researchers seek to achieve more accurate and comprehensive predictions of pollutant concentrations in both time and space.

Another noteworthy approach proposed by Zhang et al. [18] involves the parallel random forest algorithm for air quality prediction. The random forest algorithm, a powerful ensemble learning technique, combines multiple decision trees to improve predictive accuracy. In this particular study, the researchers introduced parallel processing to the algorithm, enhancing its efficiency and scalability. This modification allows the model to handle large-scale air quality datasets more effectively.

Furthermore, Gao et al. [19] explored the feasibility of using a neural network model to predict air pollutant concentrations. Neural networks are well-known for their ability to capture complex patterns in data. In this study, the author focused on utilizing a neural network to predict pollutant concentrations based on six meteorological features and time variables. While the results demonstrated the potential of neural networks in air quality prediction, the limitation lies in the relatively small number of features considered. The model's performance could be further enhanced by incorporating a more comprehensive set of environmental factors and meteorological data. In current methods for air quality prediction exhibit diversity, with researchers investigating various combinations of models and techniques. The hybrid approaches seek to leverage both temporal and spatial information to provide more accurate and holistic predictions. Meanwhile, the application of ensemble learning and neural networks demonstrates the capability of handling complex data patterns. Nevertheless, continuous efforts are required to optimize these methods, including the integration of a broader range of features and exploration of new approaches, to advance air quality prediction capabilities and effectively address environmental and public health concerns.

Although the above methods have made some progress, they are with some limitations and are especially unsuitable for processing a significant amount of data. Their training efficiency is relatively low and lack of deeply mining temporal features. Zheng et al. [20] proposed a collaborative training framework consisting of a spatial classifier and a time classifier to provide fine-granularity air quality prediction in real time using the related features as input. Hsieh et al. [21] designed an inference model based on the urban dynamic monitoring data and constructed the methodology of recommending the location of placing air quality monitoring stations by integrating the entropy minimization model. Both of the above methods were used to infer the air quality of the entire city in real time, rather than predicting the air quality for a future period.

To address the challenges posed by large-scale data in air quality forecasting, some researchers have turned to deep learning methods as a potential solution. Wang and Song [22] introduced the STE model, a fusion model that considers temporal characteristics, spatial characteristics, and weather correlations, with a specific focus on the temporal predictor. The model is based on deep LSTM (Long Short-Term Memory) networks, enabling it to capture both long-term and short-term dependencies in the data. This approach proves to be effective in learning intricate patterns and improving the accuracy of air quality predictions. Similarly, Huang and Kuo [23] proposed an LSTM-based network to forecast urban air quality. LSTM is a popular choice for capturing temporal features due to its ability to handle sequential data effectively. Both the STE model and the LSTM-based network use historical air quality data to make predictions, but they have limited consideration for prediction data with strong correlations.

In contrast, our approach in this paper leverages actual meteorological data as meteorological features and incorporates predictive data as a new feature in the prediction scheme. By doing so, we aim to enhance the model's ability to capture the impact of forecasted weather conditions on air quality. In addition, we construct statistical features to further improve the accuracy of the model by deepening the exploration of relevant data characteristics. While Wang and Song [22] also used forecasting data in their prediction work, they only treated it as historical meteorological features. In contrast, our approach fully integrates predictive data into the model to enhance its predictive capabilities. By considering both historical and predictive data as essential features, we create a more comprehensive and accurate air quality forecasting model. This research contributes to the field of environmental prediction and demonstrates the potential of deep learning techniques in improving air quality forecasting.

III. Proposed Methodology

In this project proposedwork aims to develop an effective Air Quality monitoring system using machine learning techniques, specifically XG Boost and Light GBM algorithms. The primary objective is to create a predictive model capable of accurately forecasting air quality levels in various regions. To achieve this, historical air quality data containing pollutant concentrations and environmental variables will be collected and preprocessed, including data cleaning and feature engineering.

Data Pre-Processing

Data pre-processing is a critical step in Air Quality monitoring using machine learning techniques like XG Boost and Light GBM. It involves cleaning and transforming the raw data to prepare it for effective model training and prediction. The specific data pre-processing steps in this context may include:

Data Cleaning: Remove any missing or invalid data points from the dataset. Missing data can lead to biased or inaccurate predictions, so it is crucial to handle them appropriately. This can involve imputation techniques like filling missing values with the mean or median of the corresponding feature.

Feature Engineering: Create new features or modify existing ones to enhance the model's predictive power. For example, one can derive new features from the original data, such as calculating daily or hourly averages of pollutant concentrations, or creating lag features to capture temporal patterns.

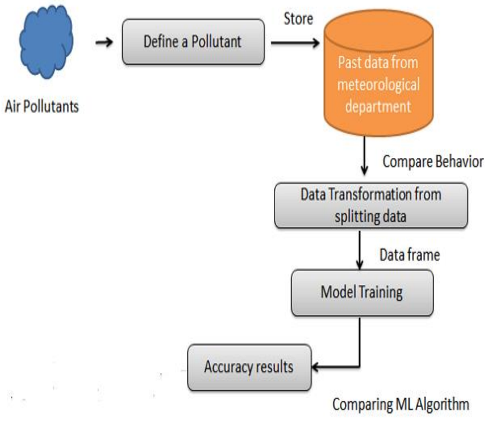

Fig.1 Proposed Block Diagram

Feature Scaling: Normalize or standardize the feature values to ensure that they are on a similar scale. This step is important, especially when using gradient boosting algorithms like XG Boost and Light GBM, as it helps the model converge faster and avoid bias towards features with larger magnitudes.

Handling Outliers: Identify and handle outliers, which are extreme values that could affect the model's performance. Outliers can be handled through various techniques like capping, winsorization, or removing them if they are genuinely erroneous.

Encoding Categorical Variables: If the dataset contains categorical variables (e.g., weather conditions), they need to be encoded into numerical values so that the machine learning models can process them. One-hot encoding or label encoding can be used for this purpose.

Train-Test Split: Divide the pre-processed dataset into training and testing sets. The training set is used to train the model, while the testing set is used to evaluate its performance and assess its ability to generalize to unseen data.

Handling Class Imbalance (if applicable): If the dataset suffers from class imbalance (e.g., when some air quality levels are rare compared to others), techniques like oversampling, undersampling, or using class weights can be employed to ensure that the model is not biased towards the majority class.

By performing these data pre-processing steps carefully, the Air Quality monitoring system can effectively leverage machine learning techniques like XG Boost and Light GBM to make accurate predictions and provide valuable insights into air pollution levels, aiding in environmental protection and public health management.

Dataset splitting

In Air Quality monitoring using machine learning techniques like XG Boost and Light GBM, dataset splitting is a crucial process to evaluate and optimize the models effectively. The dataset, containing historical air quality data and pollutant concentrations, is divided into three subsets: the training set, validation set, and test set. The training set is the largest subset and is used to train the machine learning models. During this phase, the models learn patterns and relationships from the historical data, allowing them to make accurate predictions. The training set plays a significant role in shaping the model's performance. The validation set is used for hyperparameter tuning and model selection. After training the models on the training set, they are evaluated on the validation set to assess their performance. By adjusting hyperparameters based on validation results, we can optimize the models and prevent overfitting, ensuring better generalization to unseen data.

Lastly, the test set serves as a completely unseen dataset during the model training process. It allows us to objectively assess the models' final performance and determine how well they can predict air quality levels for new, unseen data. The test set evaluation gives an unbiased estimate of the models' capabilities and their potential real-world effectiveness. By appropriately splitting the dataset into training, validation, and test sets, we can build robust and accurate models for air quality monitoring. These models enable us to predict pollutant concentrations, monitor environmental conditions, and inform decision-making to protect human health and promote sustainable environmental practices.

Model training

In Air Quality monitoring using machine learning techniques like XG Boost and Light GBM, model training is a fundamental step in developing accurate and effective predictive models. During model training, the machine learning algorithms, XG Boost and Light GBM, learn from historical air quality data and pollutant concentrations to make predictions for future air quality levels. The training process involves feeding the algorithm with the training dataset, which contains the historical air quality information. The models learn the underlying patterns and relationships between the input features (such as pollutant concentrations, weather conditions, and geographical information) and the target variable (the air quality index or pollutant levels). By iteratively adjusting the model's parameters and minimizing the prediction errors, the algorithms aim to optimize their performance.

XGBoost

XGBoost is an efficient and scalable algorithm based on tree boosting proposed by Chen & Guestrin in 2016 It is an improved version of the Gradient Boosted Decision Tree (GBDT) method. It has proven not to have its computational limitations and thus differs from the GBDT method. GBDT uses the first-order Taylor expansion, while the second-order Taylor expansion is utilized in the XGBoost's loss function. In addition, the objective function is normalized in XGBoost to alleviate the model's complexity and prevent it from overfitting. For training and tuning of our XGBoost models we have used the XGBoost python package along with it's sklearn API version. The reason we used a combination of these two packages is to use the easy-to-use and efficient grid searching capabilities for hyperparameter tuning provided by the sklearn package.

Preparation of data

The process of preparation of data for training for the XGBoost model is very similar and almost identical to the one mentioned previously for the random forest models. Except that XGBoost has a strict requirement to use only numerical features which required one-hot encoding of categorical variables in the datasets. This is further followed by transformation of the datasets in python from dataframes to an internal data structure called as DMatrix used by the XGBoost library.

Training of XGBoost models and hyperparameter tuning

XGBoost is also heavily reliant on hyperparameter tuning to give the best results. Hyperparameters were tuned using the sklearn package's grid search combined with 5-fold cross validation. A short description of these hyperparameters as described by the documentation of the packages is given below:

max_depth: refers to the maximum depth up to which a single tree is allowed to grow.

min_child_weight: refers to the minimum sum of weight needed in a child node.

gamma: refers to the minimum reduction in loss which must occur in order to partition a leaf node in the tree.

subsample: refers to the subsample ratio of training data to be used.

colsample_bytree: refers to the subsample ratio of training data to be used when growing each tree in the ensemble.

reg_alpha: refers to the L1 regularization on weights during training. A very useful parameter to control overfitting.

Light GBM

LightGBM Light Gradient Boosting Method or LightGBM is a gradient boosting framework that utilizes tree based learning algorithm. LightGBM grows trees vertically while other algorithms grow trees horizontally. The leaf with max delta loss will be chosen to grow. Leaf-wise algorithm is capable of reducing more loss than a level-wise algorithm when growing the same leaf. LightGBM can handle the large data size and takes lower memory to run.

Initialize the boosting tree as shown in equation (1):

$$f_0(x) = 0 \qquad (1)$$

Iterate m times to get the boosting tree that contains *M* decision tree:

$$f_m(x) = f_{m-1}(x) + T(x, \theta_m) \qquad (2)$$

$$f_M(x) = \sum_{m=1}^{M} T(x, \theta_m) \qquad (3)$$

where is the existing model when the *m*th iteration, is the model obtained after the *m*th iteration, is the boosting tree including *M* decision trees, is the *m*th decision tree, and θ is the parameter of the decision tree.

The proposed system aims at predicting the air quality using the LightGBM model. The model is trained with the statistical features of the historical air quality data and meteorological data collected over the past three years. By providing the weather forecast data of any particular day and pollutant data of any nearby day, we can predict the air quality of that day.

Classification

In the realm of air quality monitoring, machine learning techniques, particularly XGBoost and Light GBM algorithms, have proven to be highly effective for prediction tasks. XGBoost (Extreme Gradient Boosting) and Light GBM (Light Gradient Boosting Machine) are both popular gradient boosting algorithms that excel in handling large and complex datasets commonly encountered in air quality monitoring due to the diverse environmental variables involved. These algorithms work by combining multiple weak learners (decision trees) in an ensemble learning approach to create a robust predictive model. In air quality monitoring, prediction tasks involve predicting the air quality index, determining pollution levels, or identifying specific pollutants present in the air. XGBoost and Light GBM are well-suited for these tasks due to their ability to handle high-dimensional data and efficiently capture non-linear relationships between features.

The advantages of employing XGBoost and Light GBM include their high accuracy in prediction, making them reliable choices for critical applications that demand precision in assessing air quality for public health and environmental protection. Additionally, these algorithms offer fast computation, enabling real-time or near real-time applications where swift predictions are essential for timely interventions. Furthermore, they are capable of handling imbalanced data, which is common in air quality datasets, allowing for more balanced and accurate prediction outcomes. Finally, XGBoost and Light GBM algorithms provide valuable insights into feature importance, aiding in the identification of key factors influencing air quality, which can guide policymakers in implementing targeted measures to improve air quality and overall public health.

In XGBoost and Light GBM algorithms have emerged as powerful and versatile tools for prediction tasks in air quality monitoring. Their combination of accuracy, efficiency, and ability to handle complex data makes them invaluable in predicting air quality levels, identifying pollutants, and supporting decision-making processes to enhance air quality and safeguard the well-being of communities.

IV. Results and Discussion

DATA SPECIFICATIONS

Air quality dataset:

It contains the air quality data of the 35 air quality monitoring stations in Beijing from 2017 to 2018. Each data item of the dataset contains the id, timestamp, PM2.5 concentration, PM10 concentration, NO2 concentration, CO, O3, SO2 concentration respectively measured at the air quality monitoring stations. Table 3 provides the statistical description of the leading indicators in the air quality dataset, including the maximum, minimum, and average values of the contained data. We found that the PM2.5 concentration varies from 2 to 1004 (μg/m 3), and the maximum concentration value has exceeded the upper limit of the severe air pollution range. So it is necessary for us to employ data cleaning treatment. The PM2.5 concentration trend curve of the *aotizhongxin_aq* monitoring station within one year. From the variation of the PM2.5 concentration emerges a certain periodicity and trend and the PM2.5 concentration in winter was higher than that in summer. We speculate that the condition may be due to the burning of coal during the urban heating in winter.

TABLE 1. Comparison among models.

Model	SMAPE	RMSE	MAE
adaboost	0.50484466	38.82536479	32.95747431
xgboost	0.44070702	36.09477437	29.05448398
GBDT	0.44289675	35.30601103	29.26376280
LightGBM	0.43294860	34.87113829	28.43595921
Xgboost+ LightGBM	0.42982921	33.89227721	26.68245534

In this section, we visualize the experimental results. The fitting curve in Figure 2 shows the predicted fitting effect of the proposed model on the test set, and Figure 3 shows the scattering of the actual and predicted values. Through observing Figure 3, we find that the trend of the prediction curve is basically consistent with the trend of the actual value curve and the predicted value has a positive linear relationship with the real value. The slope of the regression curve is 1.07, which proves that the proposed model has a good fitting performance. Table 1 lists the results of all models under three evaluation indicators, in which we highlight the best results for each evaluation criterion in the test set in bold.

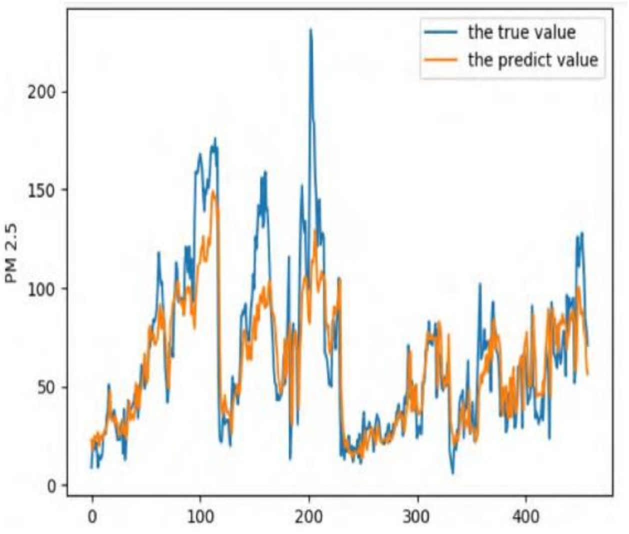

FIGURE 2. Fitting curve.

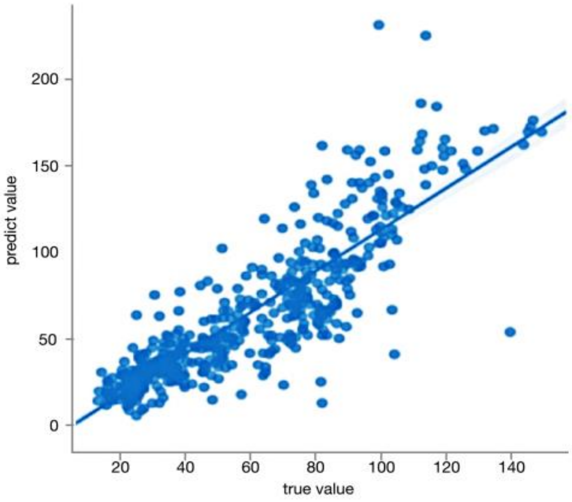

FIGURE 3. Scatter plot of actual values and prediction values.

By observing and analyzing the experimental results, we found and summarized as follows. Compared with the boost method, the neural network model shows the worst TABLE 1. Comparison among models. The reason may be that the network model is more suitable for automatically mining the features of the original dataset with the neural network and the tree model is more suitable for the tree model to dig the deep features based on data exploration under high dimensional training datasets. Among the boost methods, Adaboost showed the worst performance. The difference between GBRT and XGboost was not obvious under the three evaluation indicators. LightGBM showed the best performance according to the three evaluation indicators.

LightGBM is a machine learning algorithm that is well-suited for processing high-dimensional big data. It is a histogram-based algorithm that supports parallel learning, which makes it faster and more accurate than other boost algorithms. To verify the contribution of incorporating predictive data into the prediction process, the authors of the passage compared the performance of a model that was trained on historical data only to the performance of a model that was trained on a dataset that included both historical data and predictive data. The experimental results showed that the model that was trained on the dataset that included both historical data and predictive data had significantly better performance than the model that was trained on historical data only. The SMAPE score of the model that included predictive data was 0.07 points lower than the SMAPE score of the model that did not include predictive data. The RMSE score of the model that included predictive data was 1.0 points lower than the RMSE score of the model that did not include predictive data. And the MAE score of the model that included predictive data was 0.25 points lower than the MAE score of the model that did not include predictive data. These results suggest that predictive data can significantly improve the accuracy of machine learning models. The predictive data can help the model to learn the patterns in the data that are relevant to future predictions.

V. CONCLUSION

The quality of the air we breathe is heavily influenced by various components, including gases and particulate matter. These pollutants have a detrimental impact on air quality and can lead to serious health problems when continuously inhaled. Fortunately, with the implementation of air quality monitoring systems, we can detect the presence of these harmful substances and monitor air quality levels. This invaluable data allows us to take informed and sensible measures to improve air quality, leading to increased productivity and a reduction in health issues caused by air pollution. One of the key factors enabling these advancements is the utilization of machine learning in building prediction models. Machine learning algorithms have proven to be highly reliable and consistent in analyzing air quality data. With the aid of advanced technology and precise sensors, data collection has become simpler and more accurate than ever before. Machine learning algorithms are capable of handling the complex and extensive environmental data, allowing us to make accurate and efficient predictions. These predictions play a crucial role in understanding air quality dynamics and guiding us in implementing effective measures to protect public health and create a healthier environment.

Air quality monitoring and prediction through machine learning are powerful tools in combating air pollution. By identifying pollutants, monitoring air quality, and making informed decisions, we can elevate production levels while reducing health problems associated with poor air quality. Machine learning algorithms serve as indispensable tools in processing vast environmental data, providing us with valuable insights to build a cleaner, safer, and healthier future.

REFERENCES

[1] J. Huang et al., ‘‘A crowdsource-based sensing system for monitoring finegrained air quality in urban environments,’’ IEEE Internet Things J., to be published.

[2] X. Li, L. Peng, Y. Hu, J. Shao, and T. Chi, ‘‘Deep learning architecture for air quality predictions,’’ Environ. Sci. Pollut. Res., vol. 23, no. 22, pp. 22408–22417, 2016.

[3] Q. Zhou, H. Jiang, J. Wang, and J. Zhou, ‘‘A hybrid model for PM2.5 forecasting based on ensemble empirical mode decomposition and a general regression neural network,’’ Sci. Total Environ., vol. 496, pp. 264–274, Oct. 2014.

[4] S. Hochreiter and J. Schmidhuber, ‘‘Long short-term memory,’’ Neural Comput., vol. 9, no. 8, pp. 1735–1780, 1997.

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

[6] D. Zhu, C. Cai, T. Yang, and X. Zhou, ‘‘A machine learning approach for air quality prediction: Model regularization and optimization,’’ Big Data Cogn. Comput., vol. 2, no. 1, p. 5, 2018.

[7] D. Wang, S. Wei, H. Luo, C. Yue, and O. Grunder, ‘‘A novel hybrid model for air quality index forecasting based on two-phase decomposition technique and modified extreme learning machine,’’ Sci. Total Environ., vol. 580, pp. 719–733, Feb. 2017.

[8] G. Ke et al., ‘‘LightGBM: A highly efficient gradient boosting decision tree,’’ in Proc. Adv. Neural Inf. Process. Syst., 2017, pp. 3149–3157.

[9] W. Sun et al., ‘‘Intelligent in-vehicle air quality management: A smart mobility application dealing with air pollution in the traffic,’’ in Proc. 23rd World Congr. Intell. Transp. Syst., Melbourne, Victoria, Australia, 2015, pp. 1–12.

[10] C. Ma et al., ‘‘Reducing air pollution exposure in a road trip,’’ in Proc. 24th World Congr. Intell. Transp. Syst., Montreal, Canada, 2017, pp. 1–12.

[11] 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.

[12] W. Sun et al., ‘‘Moving object map analytics: A framework enabling contextual spatial-temporal analytics of Internet of Things applications,’’ in Proc. IEEE Int. Conf. Service Oper. Logistics, Inform. (SOLI), Jul. 2016, pp. 101–106.

[13] S. S. Roy, C. Pratyush, and C. Barna, ‘‘Predicting ozone layer concentration using multivariate adaptive regression splines, random forest and classification and regression tree,’’ in Proc. Int. Workshop Soft Comput. Appl., 2016, pp. 140–152.

[14] J. C. Chang and S. R. Hanna, ‘‘Air quality model performance evaluation,’’ Meteorol. Atmos. Phys., vol. 87, nos. 1–3, pp. 167–196, 2004.

[15] E. Meijering, ‘‘A chronology of interpolation: From ancient astronomy to modern signal and image processing,’’ Proc. IEEE, vol. 90, no. 3, pp. 319–342, Mar. 2002. [16] 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.

[17] Y. Zheng et al., ‘‘Forecasting fine-grained air quality based on big data,’’ in Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining. New York, NY, USA: ACM, 2015, pp. 2267–2276.

[18] C. Zhang and D. Yuan, ‘‘Fast fine-grained air quality index level prediction using random forest algorithm on cluster computing of spark,’’ in Proc. IEEE 12th Int. Conf. Ubiquitous Intell. Comput. IEEE 12th Int. Conf. Autonomic Trusted Comput. IEEE 15th Int. Conf. Scalable Comput. Commun. Associated Workshops, Aug. 2015, pp. 929–934.

[19] M. Gao, L. Yin, and J. Ning, ‘‘Artificial neural network model for ozone concentration estimation and Monte Carlo analysis,’’ Atmos. Environ., vol. 184, pp. 129–139, Jul. 2018.

[20] Y. Zheng, F. Liu, and H.-P. Hsieh, ‘‘U-air: When urban air quality inference meets big data,’’ in Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining. New York, NY, USA: ACM, 2013, pp. 1436–1444.

[21] H.-P. Hsieh, S.-D. Lin, and Y. Zheng, ‘‘Inferring air quality for station location recommendation based on

urban big data,'' in Proc. 21th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining. New York, NY, USA: ACM, 2015, pp. 437–446.

[22] J. Wang and G. Song, ''A deep spatial-temporal ensemble model for air quality prediction,'' Neurocomputing, vol. 314, pp. 198–206, Nov. 2018.

[23] C. J. Huang and P.-H. Kuo, ''A deep CNN-LSTM model for particulate matter (PM2.5) forecasting in smart cities,'' Sensors, vol. 18, no. 7, p. 2220, 2018.

[24] A. Shishegaran, M. Saeedi, A. Kumar, and H. Ghiasinejad, "Prediction of air quality in Tehran by developing the nonlinear ensemble model," Journal of Cleaner Production, vol. 259, Article ID 120825, 2020.

[25] P. Bhalgat, S. Pitale, and S. Bhoite, "Air quality prediction using machine learning algorithms," International Journal of Computer Applications Technology and Research, vol. 8, pp. 367–370, 2019.

[26] Tanisha Madan, Shrddha Sagar, Deepali Virmani, " Air Quality Prediction using Machine Learning Algorithms", 2020, IEEE.

[27] C. R. Aditya, C. R. Deshmukh, N. D K, P. Gandhi, and V. astu, "Detection and prediction of air pollution using machine learning models," International Journal of Engineering Trends and Technology, vol. 59, no. 4, pp. 204–207, 2018.

[28] J. Kleine Deters, R. Zalakeviciute, M. Gonzalez, and Y. Rybarczyk, "Modeling PM2. 5 urban pollution using machine learning and selected meteorological parameters," Journal of Electrical and Computer Engineering, vol. 2017, Article ID 5106045, 14 pages, 2017.

[29] A. G. Soundari, J. Gnana, and A. C. Akshaya, "Indian air quality prediction and analysis using machine learning," International Journal of Applied Engineering Research, vol. 14, p. 11, 2019.

[30]S. Halsana, "Air quality prediction model using supervised machine learning algorithms," International Journal of Scientific Research in Computer Science, Engineering and Information Technology, vol. 8, pp. 190–201, 2020.