International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Air Quality Index Forecasting via Genetic Algorithm-Based Improved Extreme Learning Machine

M. Rakshitha[1], Dr.MD.Adam Baba[2], B.Ramsai[3], M.Ram Yashvanth[4], R.Ramcharan[5], S.Ramesh[6]

[CSE]-AI & ML MRUH.

ABSTRACT:

Air quality has always been one of the most important environmental concerns for the general public and society. Using machine learning algorithms for Air Quality Index (AQI) prediction is helpful for the analysis of future air quality trends from a macro perspective. When conventionally using a single machine learning model to predict air quality, it is challenging to achieve a good prediction outcome under various AQI fluctuation trends. In order to effectively address this problem, a genetic algorithm-based improved extreme learning machine (GA-KELM) prediction method is enhanced. First, a kernel method is introduced to produce the kernel matrix which replaces the output matrix of the hidden layer. To address the issue of the conventional limit learning machine where the number of hidden nodes and the random generation of thresholds and weights lead to the degradation of the network learning ability, a genetic algorithm is then used to optimize the number of hidden nodes and layers of the kernel limit learning machine. The thresholds, the weights, and the root mean square error are used to define the fitness function. Finally, the least squares method is applied to compute the output weights of the model. Genetic algorithms are able to find the optimal solution in the search space and gradually improve the performance of the model through an iterative optimization process. In order to verify the predictive ability of GA-KELM, based on the collected basic data of long-term air quality forecast at a monitoring point in a city in China, the optimized kernel extreme learning machine is applied to predict air quality (SO2, NO2, PM10, CO, O3, PM2.5 concentration and AQI), with comparative experiments based CMAQ (Community Multiscale Air Quality), SVM (Support Vector Machines) and DBN-BP (Deep Belief Networks with Back-Propagation). The results show that the proposed model trains faster and makes more accurate predictions. As extension we have experimented with BI-LSTM algorithm which will optimize features weight in both forward and backward direction.

1.Introduction

Air pollution is a prevalent environmental problem in the twenty-first century. In light of the rapid industrialization and urbanization, air pollution is getting worse, which greatly affects our living environment and health. Li et al. came to the conclusion that outdoor physical activity poses numerous health risks due to ambient air pollution in China. According to the Chinese Ambient Air Quality Standards (GB3095-2012), there are six conventional air pollutants used to measure air quality: sulfur dioxide (SO2), nitrogen dioxide (NO2), particulate matter with a particle size less than 10 microns (PM10), particulate matter with a particle size less than 2.5 microns (PM2.5), ozone (O3), and carbon monoxide (CO). These pollutants have adverse effects on human health. The International Energy Agency estimates that air pollution causes 6.5 million premature deaths per year, while long-term exposure to pollutants, such as fine particles (e.g.,PM2.5) or traffic-related pollutants, is linked to higher rates of lung cancer, coronary heart disease, and other illnesses. Therefore, studies on air quality prediction are particularly important and are considered a key factor for environmental protection. In order to more comprehensively assess the health effects of air pollution, numerous air quality monitoring stations have been set up in major cities. Air quality predictions can be made based on the data collected from these stations. Air quality monitoring, modeling, and accurate predictions are important for having a clear understanding of future

2.Literature Review

Air pollution is a prevalent environmental problem in the twenty-first century. In light of the rapid industrialization and urbanization, air pollution is getting worse, which greatly affects our living environment and health. Li et al. came to the conclusion that outdoor physical activity poses numerous health risks due to ambient air pollution in China. According to the Chinese Ambient Air Quality Standards (GB3095-2012), there are six conventional air pollutants used to measure air quality: sulfur dioxide (SO2), nitrogen dioxide (NO2), particulate matter with a particle size less than 10 microns (PM10), particulate matter with a particle size less than 2.5 microns (PM2.5), ozone (O3), and carbon monoxide (CO). These pollutants have adverse effects on human health. The International Energy Agency estimates that air pollution causes 6.5 million premature deaths per year, while long-term exposure to pollutants, such as fine particles (e.g.,PM2.5) or traffic-related pollutants, is linked to higher rates of lung cancer, coronary heart disease, and other illnesses. Therefore, studies on air quality prediction are particularly important and are considered a key factor for environmental protection. In order to more comprehensively assess the health effects of air pollution, numerous air quality monitoring stations have been set up in major cities. Air quality predictions can be made based on the data collected from these stations. Air quality monitoring, modeling, and accurate predictions are important for having a clear understanding of future pollution levels and their associated health risks.

Several studies have explored the application of genetic algorithms (GAs) to enhance air quality index (AQI) forecasting models. One notable approach involves integrating GAs with fuzzy time series models to improve air quality management. In the study "Air Quality Management Using Genetic Algorithm-Based Heuristic Fuzzy Time Series Model," researchers developed a hybrid adaptive time-variant fuzzy time series model optimized by GAs. This model was applied to predict AQI in two major Indian cities, demonstrating improved accuracy over existing models.
Emerald
Another significant study, "Predicting Air Quality Index Using an Improved Extreme Learning Machine Based on Genetic Algorithms," addresses the challenges of AQI prediction due to varying patterns in AQI fluctuations. The researchers developed an extreme learning machine (ELM) prediction technique enhanced by GAs (GA-KELM). By optimizing the number of hidden nodes and layers, and applying a kernel method, the GA-KELM model demonstrated faster training times and more accurate air quality predictions compared to models like CMAQ, SVM, and DBN-BP.
IJERS
Further extending this approach, the study "Using Deep Learning and an Enhanced Extreme Learning Machine Based on Genetic Algorithms to Forecast Air Quality Index" combines deep learning techniques with the GA-KELM model. This hybrid approach aims to capture complex, nonlinear relationships in air quality data, resulting in improved prediction accuracy and efficiency.
Computer Science Journals
These studies collectively highlight the potential of integrating genetic algorithms with various modeling techniques to enhance AQI forecasting, thereby contributing to more effective air quality management strategies

Genetic Algorithm-Based Heuristic Fuzzy Time series

Model: This study introduces a heuristic fuzzy time series model optimized with genetic algorithms to predict AQI. The model was tested on real-time air pollution data from two Indian cities, demonstrating improved accuracy over existing models.

ResearchGate

Optimized Machine Learning Model Using Grey Wolf Optimization: Researchers proposed a regression model combining the Grey Wolf Optimization algorithm with decision tree regression to predict AQI in major Indian cities. The model effectively extracted optimal features from historical data, enhancing prediction accuracy.

Nature

Bidirectional Recurrent Neural Networks for Daily AQI Forecasting: This research focused on accurately forecasting daily AQI using bidirectional recurrent neural networks, addressing the challenges posed by the complex and nonlinear nature of air quality data.

ScienceDirect

Improved Extreme Learning Machine Based on Genetic Algorithms: A study presented a comparative analysis of models including Support Vector Regression (SVR), Genetic Algorithm-Enhanced Extreme Learning Machine (GA-KELM), and Deep Belief Network with Back-Propagation (DBN-BP). The integration of Bidirectional Long Short-Term Memory (BiLSTM) was proposed to further enhance prediction accuracy, with results indicating that BiLSTM outperformed existing models.

IJGST

Enhancing Extreme Learning Machines with Genetic Algorithms: This study proposed a hybrid model that combines genetic algorithms with extreme learning machines to improve air quality forecasting accuracy. The model demonstrated superior performance in predicting AQI, contributing to informed decision-making for pollution control strategies and public health interventions.

3.Methodology

MODULES:

Data loading: using this module we are going to import the dataset.
Data Processing: Using the module we will explore the data.
Splitting data into train & test: using this module data will be divided into train & test
Model generation: Model building - SVR - GA-KELM - DBN-BP(NN with Back-Propagation) - BiLSTM. Algorithms accuracy calculated
User signup & login: Using this module will get registration and login
User input: Using this module will give input for prediction
Prediction: final predicted displayed
Note: Extension
In the base paper the author mentioned to use different deep and machine learning model for analysis the Air Quality dataset, like DBN, SVR and GA-KELM,

However, we can further enhance the performance by exploring other techniques such as BiLSTM which got less rmse and mse value compare with proposed model,

Algorithms:

SVR: Support Vector Regression (SVR) is a type of machine learning algorithm used for regression analysis. The goal of SVR is to find a function that approximates the relationship between the input variables and a continuous target variable, while minimizing the prediction error.

GA-KELM: GA-KELM (Genetic Algorithm-based Kernel Extreme Learning Machine): A predictive algorithm for air quality forecasting. It incorporates a kernel method to generate a kernel matrix, replacing the hidden layer output matrix. A genetic algorithm optimizes the number of hidden nodes and layers, considering thresholds, weights, and root mean square error in the fitness function. Output weights are computed using the least squares method, enabling iterative optimization for improved model performance.

DBN-BP(NN with Back-Propagation): DBN-BP (Deep Belief Networks with Back-Propagation): A neural network algorithm for air quality prediction. Deep Belief Networks are trained layer-wise, and Back-Propagation fine-tunes the model. It combines unsupervised pre-training with supervised learning, optimizing weights iteratively. This hybrid approach enhances the model's ability to capture complex patterns in air quality data, leading to more accurate predictions.

4.Design

3.1System Architecture:

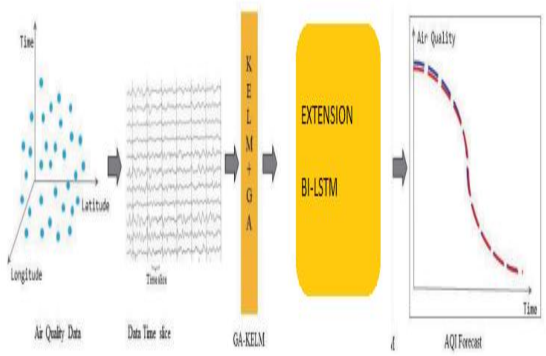
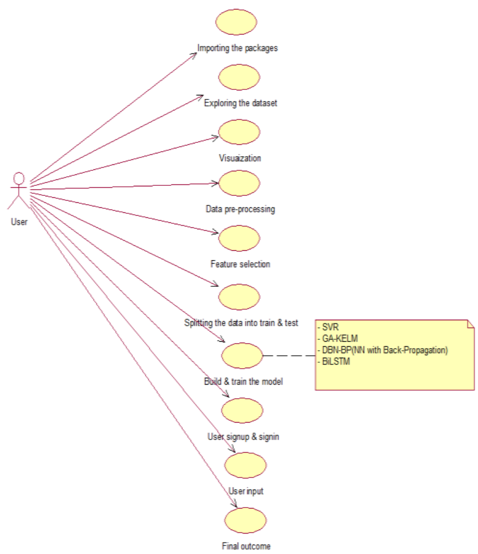


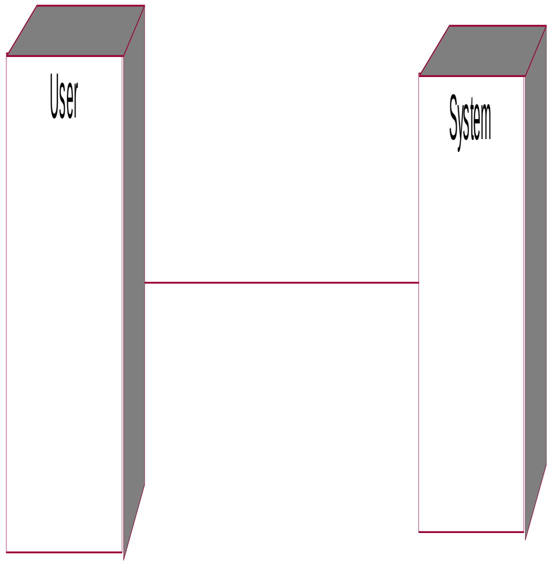
Fig.5.1.1 System architecture

3.2 UML Diagrams

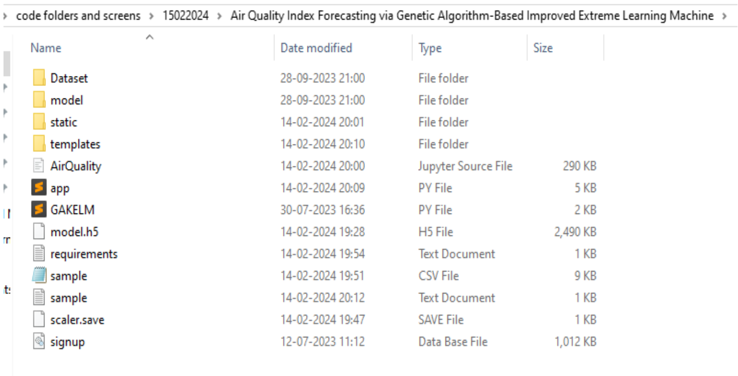


3.3Deployment diagram:

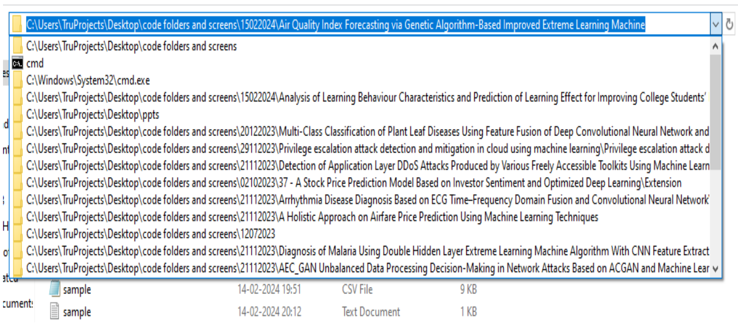
The deployment diagram captures the configuration of the runtime elements of the application. This diagram is by far most useful when a system is built and ready to be deploy.



5.Results

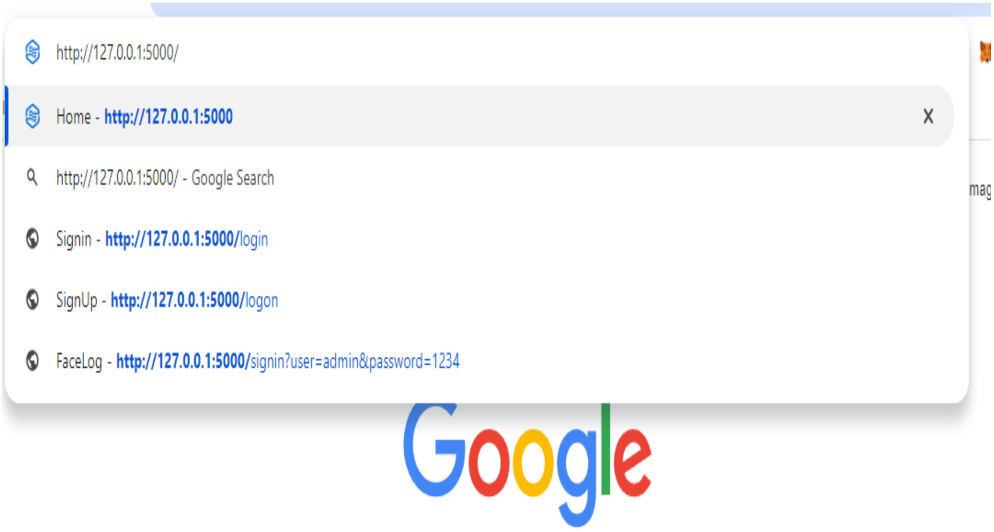


Step 1

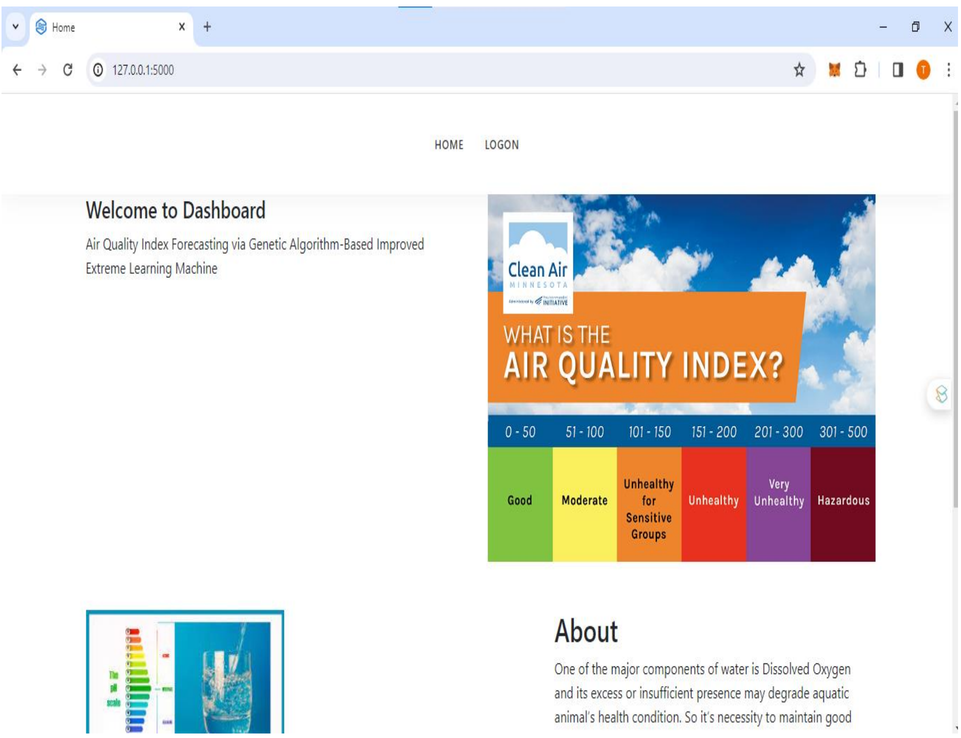


Step 2

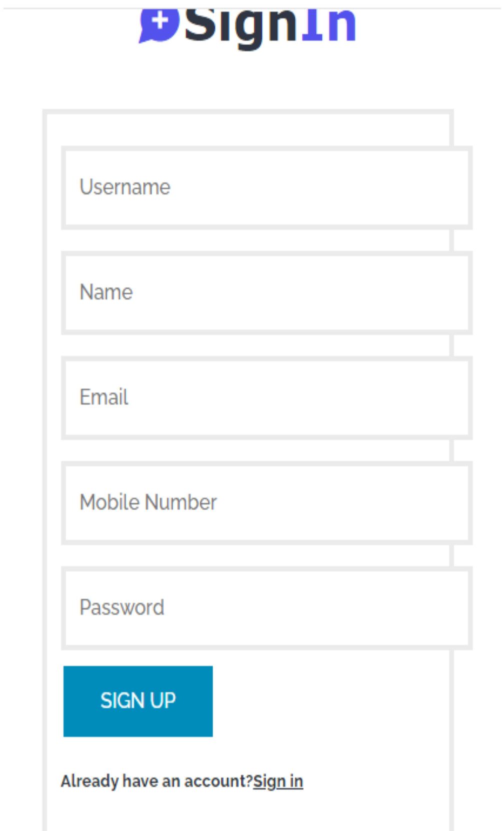
Step 3



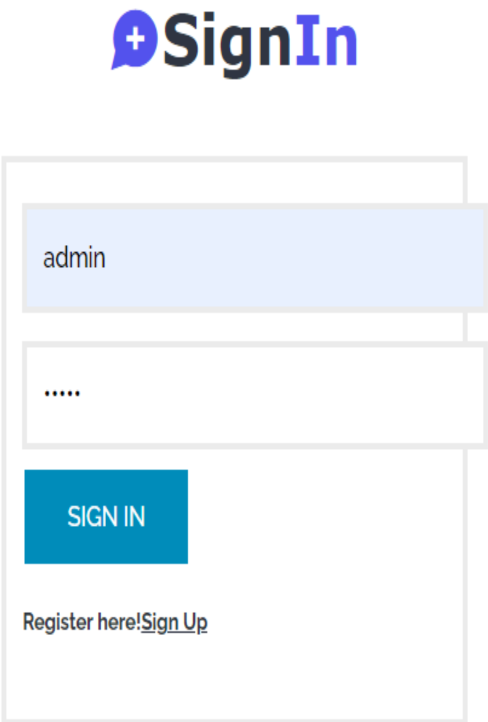
Step 4



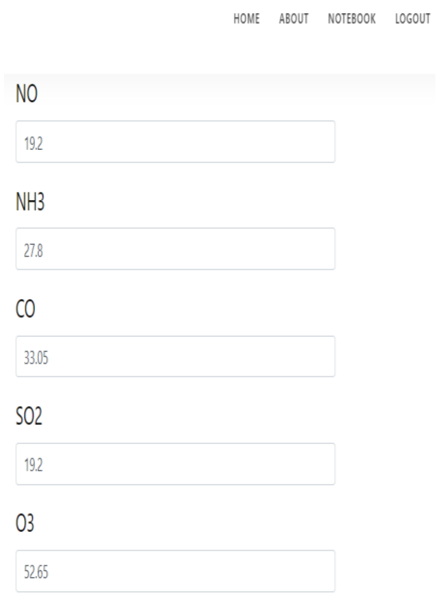
Step 5


SignIn
Username
Name
Email
Mobile Number
Password
SIGN UP
Already have an account?Sign in

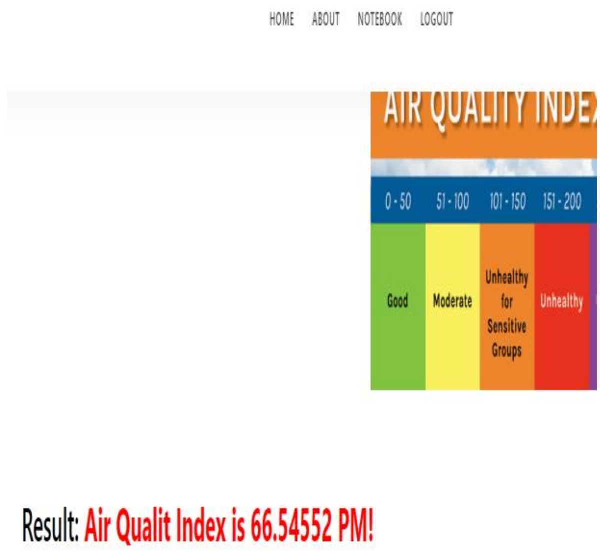
Step 6


SignIn
admin
SIGN IN
Register here!Sign Up

Step 7



Step 8



Step 9

6.Conclusion

The economic development achieved by the country through rapid urbanization is polluting the environment in an alarming way and putting people's lives in danger. Therefore, a correct analysis and accurate prediction of air quality remains a primary condition to achieve the objective of sustainable development. This paper focuses on the problem of prediction model design, and investigates the problems related to the optimization of the model parameters. A GA-KELM model is designed, implemented, and tested. It is experimentally proven to be more efficient than the classical shallow learning and can effectively explore and learn the interdependence of multivariate air quality correlation time series such as temperature, humidity, wind speed, SO2, and PM10. Therefore, the GA-KELM model developed in this study can be used to provide valuable support to vulnerable groups and trigger early warning of adverse air quality events. However, there are still areas for further investigation and improvement. In recent years, numerous advanced algorithms and optimization methods based on genetic algorithms and population intelligence have emerged.

Therefore, future research should explore the underlying significance and value of combinatorial intelligence optimization algorithms such as the Limit Learning Machine. Additionally, we acknowledge the need to address the issue of manually setting the number of hidden layer nodes in the optimal Limit Learning Machine. Although the Dynamic Extreme Learning Machine (DELM) algorithm offers adaptive determination of hidden layer nodes without human intervention, further work should be dedicated to this aspect. Moreover, to enhance the accuracy and validity of air quality measurement and assessment, it is crucial to integrate pollutant emission factors and meteorological factors into the evaluation system.

This integration will enable a more precise and comprehensive evaluation of air quality. In conclusion, our study highlights the significance of the GA-KELM model in predicting air quality. We have addressed the optimization challenges and demonstrated its superiority over traditional methods. However, there is still room for improvement and further research. Future studies should delve into advanced optimization algorithms based on genetic algorithms and population intelligence, explore the potential of the Limit Learning Machine, and strive for adaptive determination of hidden layer nodes. Furthermore, the integration of pollutant emission factors and meteorological factors into the evaluation system will advance the accuracy and reliability of air quality measurement and assessment.
As extension we have experimented with BI-LSTM algorithm which will optimize features weight in both forward and backward direction. BI-LSTM will optimized features till no more optimizations are possible so prediction accuracy will automatically get increased and error MSE rate will get decrease.

7. Future Work

1. Advancements in Model Optimization and Hybrid AI Approaches

The current implementation of GA-ELM for AQI forecasting can be further improved by integrating it with other metaheuristic algorithms like Particle Swarm Optimization (PSO), Ant Colony Optimization (ACO), Grey Wolf Optimizer (GWO), and Artificial Bee Colony (ABC). These techniques can help refine the selection of input weights and biases in the ELM model, leading to improved accuracy and generalization. Additionally, multi-objective optimization methods such as NSGA-II (Non-dominated Sorting Genetic Algorithm II) can balance prediction accuracy and computational efficiency, which is crucial for large-scale implementations.

Another promising direction is the hybridization of GA-ELM with deep learning models such as Long Short-Term Memory (LSTM) networks and Gated Recurrent Units (GRUs). These deep learning architectures are highly effective in capturing spatiotemporal dependencies in time-series AQI data. A GA-ELM-LSTM hybrid model could leverage the fast training speed of ELM while utilizing LSTM's ability to handle long-term dependencies, leading to superior forecasting accuracy. Additionally, adaptive learning strategies where mutation and crossover rates evolve dynamically based on model performance can further enhance the robustness of GA-ELM.

2. Real-Time and Large-Scale AQI Forecasting

One major limitation of traditional AQI forecasting models is their inability to operate in real-time due to computational constraints. Future advancements can focus on integrating GA-ELM with edge computing and Internet of Things (IoT) networks. Deploying AI models on IoT-enabled air quality monitoring sensors will enable real-time AQI predictions without relying on centralized cloud processing. This decentralized approach reduces latency and enhances the efficiency of smart city infrastructures.

Cloud-based AI solutions, such as those using Google Cloud AI, AWS Machine Learning, and Microsoft Azure AI, can facilitate large-scale AQI forecasting across multiple cities and regions. The use of federated learning (FL) can further enhance privacy and scalability by allowing multiple air quality monitoring stations to train models collaboratively without sharing raw data. This technique is particularly useful for protecting sensitive environmental and health-related information.

Another promising research direction is high-resolution spatial AQI mapping. Many urban areas lack sufficient monitoring stations, leading to gaps in AQI data. By integrating GA-ELM with Geographic Information System (GIS) technologies, researchers can develop models that interpolate AQI levels across regions with limited sensor coverage. This will provide better air quality insights for urban planning and environmental policy-making.

8.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.