IJARR, 10(4), 2025; 89-93

International Journal of Advanced Research and Review

www.ijarr.org

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

Poojitha. N [1], Rohan Keshava Rao. A, [2], Sri Sai Varsha. P[3], Mr. P.Krishna Reddy[4]
[123]UG Scholar, Dept. of AI&DS, St. Martin's Engineering College, Secuenderabad, Telangana, India – 500100
[4] Assistant Professor, , Dept. of AI&DS, St. Martin's Engineering College, Secuenderabad, Telangana, India – 500100
poojithanarayanagari09@gmail.com

Abstract:

Air quality has always been one of the most important environmental concerns for the 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. 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 can 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 (SO_{2} ,NO_{2} , PM_{10} , CO , O_{3} , PM_{2.5} 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.

Keywords: Air Quality Index (AQI), Machine Learning, Air Pollution Prediction, Extreme Learning Machine (ELM),Kernel Extreme Learning Machine (KELM)Genetic Algorithm (GA),GA-KELM.

1. INTRODUCTION

Air pollution is aprevalentenvironmentalprobleminthetwenty-firstcentury.Considering the rapid industrialization and urbanization, air pollution is getting worse, which greatly affects our living environment and health. Li et al. concluded that outdoor physical activity poses numerous health risks due to ambient air pollution in China. According to the Chinese Ambient Air Quality Standard identify six key pollutants: sulphur dioxide (SO2), nitrogen dioxide (NO2), particulate matter (PM10 and PM2.5), ozone (O3), and carbon monoxide (CO). These pollutants contribute to severe health problems, including lung cancer and coronary heart disease. The International Energy Agency estimates that air pollution causes 6.5 million premature deaths annually.

Air quality monitoring stations collect data for pollution prediction, aiding environmental protection. Machine learning algorithms are increasingly used for air quality prediction, though traditional models face challenges such as slow learning and complex training. Wu et al. developed an optimal-hybrid model for AQI prediction using all six pollutants, while Huang et al. introduced an extreme learning machine (ELM) algorithm to improve prediction accuracy and training speed. However, ELM's random selection of hidden layer parameters affects prediction stability.

To address this, a GA-based improved extreme learning machine (GA-KELM) is proposed. This model optimizes ELM's hidden layer nodes, thresholds, and weights using a genetic algorithm (GA) with root mean square error (RMSE) as the fitness function. Continuous coding updates parameters, forming a hierarchical control structure. GA-KELM is tested against CMAQ, SVR, and DBN-BP models, showing superior prediction accuracy.

The study's main contributions include improving ELM's activation function, enhancing prediction accuracy with GA optimization, and performing correlation analysis of air quality factors. Experimental results on three real-world datasets confirm GA-KELM's effectiveness, making it a reliable method for air quality prediction. Future research will focus on further refining the model.

2. LITERATURE SURVEY

Rapid socio-economic development and urbanization have resulted in serious deterioration in air-quality in many world cities, including Beijing, China.T hisstudy attempts to examinetheeffectivenessofairpollutioncontrolregulationsimplemented in Beijing during 2008–2019 through a data-driven regulatory intervention analysis. Our proposed Bayesian deep learning model utilizes proxy data including Aerosol Optical Depth (AOD) and meteorology as well as socio-economic data, while accounting for confounding effects via propensity score estimation. Our resultsshow that air pollution control regulatory measures implemented in China and Beijing during 2008–2019 reduced PM2.5 pollution in Beijing by 11 % on average.Afterthe introduction of the Action Plan for Clean Air in China and Beijing in late 2013, as compared to the hypothetical PM2.5 concentration (without any regulatory interventions), the estimated PM2.5 reduction increased dramatically from 15 % in 2015 to 44 % in 2018. Our results suggest that Beijing's air quality has improved gradually over the past decade, though the annual PM2.5 pollution still exceeds the WHO threshold. In this regard, the air pollution control regulations introduced in Beijing and China tend tobecomemoreeffectiveafter2015,suggestinga2-yeartime lag before the stringent air pollution control regulations starting from 2013 takesany strong positive effects. Moreover, as compared totheairpollutioncontrolregulations introduced before 2013, newly introduced policy-making governance, which couples the policy-makings of the local jurisdictions with that of the central government,and the new policy measures that tackle the vested interests of the local stakeholders in Beijing and its nearby cities, alongside with the stringent local and national air pollution control regulations and plans, should help reduce air pollution andpromote healthy living in Beijing over the longer term.

Air pollution is defined as a phenomenon harmful to the ecological system and the normal conditions of human existence and development when some substances in the atmosphere exceed a certain concentration. In the face of increasingly serious environmental pollution problems, scholars have conducted a significant quantity of related research, and in those studies, the forecasting of air pollution has been ofparamount importance. As a precaution, the air pollution forecast is the basis for taking effective pollution control measures, and accurate forecasting of air pollution has become an important task. Extensive research indicates that the methods of air pollution forecasting can be broadly divided into three classical categories: statistical forecasting methods, artificial intelligence methods, and numerical forecasting methods. More recently, some hybrid models have been proposed, which can improve the forecast accuracy. To provide a clear perspective on air pollution forecasting, this study reviews the theory and application of those forecasting models. In addition, based on a comparison of different forecasting methods, the advantages and disadvantages of some methods of forecasting are also provided. This study aims to provide an overview of air pollution forecasting methods for easy access and reference by researchers, which will be helpful in further studies.

To the best of our knowledge, there are a few research on air-handwriting character-level writer identification only employing acceleration andangularvelocity data. In this paper, we propose a deep learning approach to writer identification only using inertial sensor data of air-handwriting. We separate different representations of degree of freedom (Doff) of air-handwriting to extract local dependency and interrelationship in different CNNs separately. Experiments on a public dataset achieve an average good performance with out any extrahand-designed feature extractions.

Particulate matter (PM_{10}) is the pollutant causing exceedances of ambient air quality thresholds, and the key indicator of air quality index in Brunei Darussalam for haze related episodes caused by the recurrent biomass fires in Southeast Asia.Thepresent study aims at providing suitable forecasts for PM_{10} exceedances to aid in health advisory during haze episodes at the four administrative districts of the country. A framework based on random forests (RFs), genetic algorithm (GA) and back propagation neural networks (BPNN) computational intelligence techniques hasbeen proposed in which the final prediction is made by the BPNN model. A hybrid combination of GA and RFs is initially applied to determine optimal set of inputsfrom the initial data sets of largely available meteorological, persistency of high pollutionlevels,short-andlong-termvariationsofemissionsratesparameters.Theinputs selection procedure does not depend on the back propagation training algorithm.Thenumericalresultspresentedinthispapershowthattheproposedmodel not only produced satisfactory forecasts but also consistently performed better via several statistical performance indicators when compared with the standard BPNNand GA optimization based on back propagation training algorithm. The model also showed satisfactory threshold exceedances forecasts achieving for instance best true predicted rate of 0.800, false positive rate of 0.014, false alarm rate of 0.333 and success index of 0.786 at Brunei-Maura district monitoring station. Overall, the current study has profound implications on future studies to develop a real-time air quality forecasting system to support haze management.

3. PROPOSED METHODOLOGY

AQI is first introduced, the ELM and KELM algorithms are presented next, and a new GA-KELM learning method for AQI prediction is then proposed.

Air quality forecasting has been a key issue in early warning and control of urban air pollution. Its goal is to anticipate changes in the AQI value at observation points over time. The observation period, which is decided by the ground-based air-quality monitoring station, is usually set for one hour.

Furthermore, a location's air quality value is largely influenced by the weather conditions prevailing at that location. Air quality monitoring stations measure air temperature, wind direction, atmospheric pressure, relative humidity, wind speed and other meteorological parameters, as well as air pollutant concentrations. Air quality prediction is also challenging due to the rapid changes in pollutant emission and weather conditions. Numerous variables, such as wind speed, temperature, humidity, and pollutants themselves, are highly nonlinear, dynamic, and have inherent interdependencies, making it more challenging to accurately predict air quality at a specific time and place. Therefore, it is essential to figure out how to deal with these factors and exploit them from multivariate time-series data related to air quality. A typical meteorological factor sequence diagram is shown in FIGURE 1.

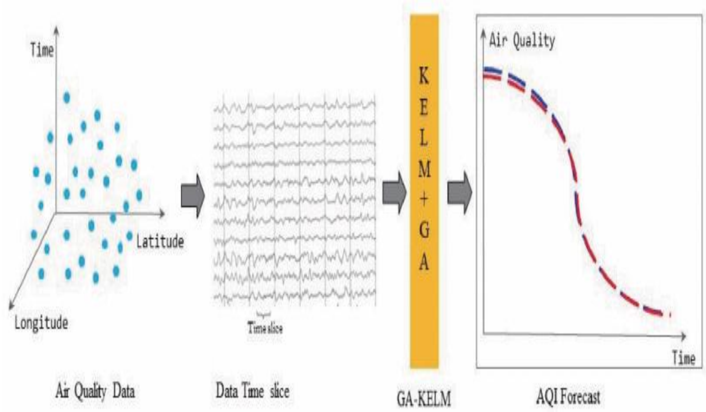

FIGURE 1

Applications:

- Environmental Monitoring – Real-time air quality forecasting for pollution control.
- Smart Cities – Integration with IoT for urban air quality management.
- Public Health – Early warnings for air pollution-related diseases.
- Industrial Compliance – Monitoring emissions to meet environmental regulations.
- Traffic Management – Predicting pollution from vehicular emissions.
- Climate Change Research – Studying long-term air quality trends.
- Energy Sector – Assessing pollution impact of fossil fuel usage.
- Disaster Management – Predicting air quality after wildfires or industrial accidents.
- Agriculture Protection – Evaluating air pollution effects on crops and soil.
- AI-Driven Policy Making – Data-driven strategies for pollution control.

Advantages:

- Higher Prediction Accuracy – GA-KELM improves air quality forecasting accuracy compared to traditional models.
- Faster Training Speed – Optimized learning process reduces computational time.
- Robust Performance – Effectively handles AQI fluctuations and nonlinear data.
- Optimized Model Parameters – Genetic Algorithm (GA) fine-tunes hidden nodes and layers for better performance.
- Better Generalization – Avoids overfitting and performs well on new data.
- Effective Feature Selection – Identifies key pollutants affecting AQI predictions.
- Scalability – Can be applied to different cities and environmental conditions.
- Reduced Human Intervention – Automated optimization minimizes manual tuning efforts.
- Versatile Applications – Useful for environmental monitoring, healthcare, traffic management, and more.
- Improved Decision-Making – Provides reliable data for policymakers and urban planners.

4. EXPERIMENTAL ANALYSIS

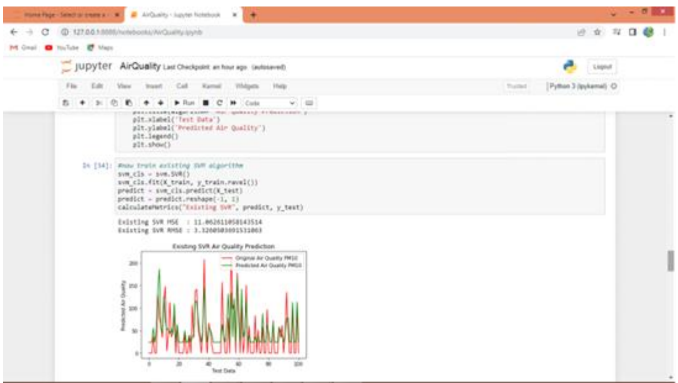

Figure 1: Performance of RMSE AND MSE.

In the above screen training SVM algorithm and then with SVM we got 11 as the MSE and in the graph x-axis represents TEST COUNT and y-axis represents air quality. Red line represents Original Test Air Quality and green line represents Predicted Air Quality and we can see both lines are closed as they are overlapping with little gaps and this gap can reduce by applying propose algorithm

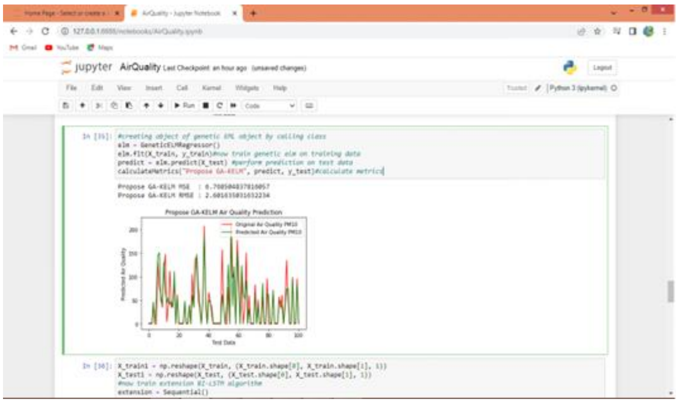

Figure 2: Performance of GA-KLEM.

In the above screen we are training to propose genetic ELM called GA-KELM and we got its MSE as 6 and in graph we can see now both lines overlap with too few gap.

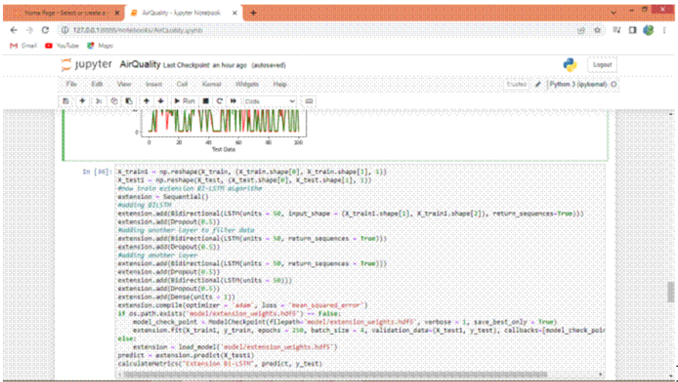

Figure 3: Code of GA-KLEM.

In above screen training extension Bi-LSTM algorithm and after executing above block will get below output.

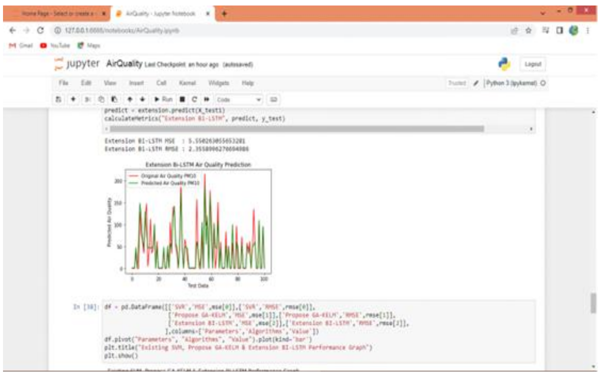

Figure 4: Performance of Bi-LSTM.

In above screen with extension Bi-LSTM we got MSE as 5% and in graph also nearly 90% test and predicted air quality lines are overlapping and in all algorithms extension Bi-LSTM has got less MSE and RMSE.

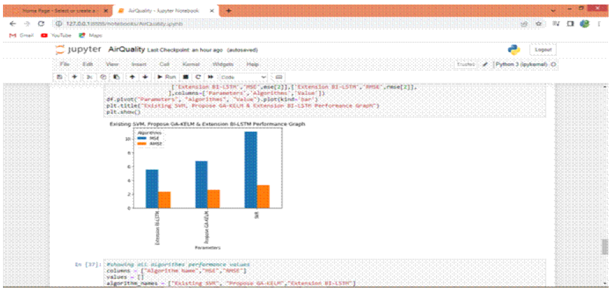

Figure 5: Comparison Graph.

In above graph x-axis represents algorithm names and y-axis represents MSE and RMSE in different colour bars and in all algorithms extension Bi-LSTM got less RMSE and MSE.

5. 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.

REFERENCES

[1] Abram, S., Chitra, P., Mahometh, R., Keserwan, S.R., 2020. Hybrid Spatio-temporal Deep Learning Framework for Particulate Matter (PM2.5) Concentration Forecasting.2020 International Conference on Innovative Trends in Information Technology (ICITIIT) 1e6.

[2] M. Jia, A. Comelily, Y. Wang, and R. S. Srinivasan, "Adopting Internet of Things for the development of smart buildings: A review of enabling technology gives and applications," Automat. Construct., vol. 101, pp. 111126, May 2019.

[3] Y. Ding and Y. Xu, "A Deep Learning Approach to Writer Identification Using Inertial Sensor Data of Air-Handwriting," IEICE Transient. & Syst., vol. E102-D, no.10, pp.2059–2063, 2019.

[4] X. Li, L. Jinn, and H. Kan, "Air pollution: A global problem needs local fixes," Nature, vol.570, pp.437–439, 2019.

[5] Y. Cheng, S. Zhang, C. Huan, M. O. Loudoun, and Z. Lin, "Optimization on fresh outdoor air ratio of air conditioning system with stratum venti-lotion for both targeted indoor air quality and maximal energy saving, "Building Environ., vol. 147, pp. 1122, Jan. 2019.

[6] C. Cosmas and R. Simcha, "Machine learning method for real-time non-invasive prediction of individual thermal preference in transient conditions," Building Environ., vol. 148, pp. 372383, Jan. 2019.

[7] Chen, R., Wang, X., Zhang, W., Zhu, X., Li, A., Yang, C., 2019. A hybrid CNN-LSTM model for typhoon formation forecasting. Geoinformatics 23, 375e396. https://doi.org/10.1007/s10707-019- 00355-0.

[8] Du, S., Li, T., Yang, Y., Hong, S.-J., 2019. Deep air quality forecasting using a hybrid deep learning framework. IEEE Trans. Knowle. Data Eng. https://doi.org/10.1109/ tkde.2019.2954510, 1-1..

[9] Feng, R., Zheng, H., Gao, H., Zhang, A., Huang, C., Zhang, J., Luo, K., Fan, J., 2019.Recurrent Neural Network and random forest for analysis and accurate forecast of atmospheric pollutants: a case study in Hangzhou, China.J. Clean. Prod.

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