Prediction of Air Quality Index using genetic programming

Chu Thi Quyen*, Ngo Thi Thanh Hoa, Nguyen Thi Cam Ngoan

Hanoi University of Industry - No 298 Cau Dien street, Bac Tu Liem district, Hanoi.
*Corresponding author: chuthiquyen_cntt@haui.edu.vn
Received 26 Aug. 2023; Revised 21 Oct. 2023; Accepted 10 Nov. 2023; Published 25 Nov. 2023.
DOI: https://doi.org/10.54939/1859-1043.j.mst.91.2023.85-95

ABSTRACT

The Air Quality Index (AQI) is a tool used to measure the impact of air pollution on health over time. In the world, air pollution has significantly increased, and machine learning techniques are used to forecast and analyze AQI. We present a new way for using GP to evolve models for AQI forecasting in this work GP can evolve more accurate AQI forecasting models than other standard machine learning algorithms, according to experimental results using datasets obtained from various stations across multiple cities in India. Furthermore, while developing AQI forecasting models, GP can automatically identify significant features, and the models developed by GP are interpretable.

Keywords: Machine Learning; Genetic Programming; AQI.

1. INTRODUCTION

Air pollution is a significant issue resulting from human activity and natural disasters, affecting the environment and affecting human health [1]. Meteorological factors like wind speed, humidity, and temperature influence air contaminants. Urbanization and industrialization contribute to air pollution, releasing pollutants like nitrogen oxide, carbon monoxide, and particulate matter. To reduce air pollution, environmental measures must be implemented, and the Air Quality Index (AQI) is used to assess air quality [2].

The raw data must be thoroughly analyzed in order to find contaminants. Machine learning algorithms ensure accurate prediction of future AQI levels so that relevant measures may be implemented. Recent research focuses more on machine learning algorithms for air quality evaluation and air pollution prediction including Neural Network, k - Nearest Neighbor, Support Vector Machine, and Random Forest [3-6].

The suggested methodology is based on Genetic Programming (GP), a soft computing search strategy that uses evolutionary algorithms. GP develops the structure and parameters of expressions at the same time, allowing formalization of the relationships between variables that best suit a reference series. This strategy is particularly beneficial in cases when the exact functional form of the solution is unknown in advance.

The research contributions of this study are significant. First, this is the study that has compared the performances of different forecasting models, such as classical and advanced techniques applied on a real-world dataset. Second, we have demonstrated the ability of GP models on AQI forecasting. Third, we have contributed to finding the variables have real meaning in the GP forecast model. Finally, this study has proven that GP models are quite good candidates for predicting the AQI of other stations.

The rest of this document is outlined as follows. In the next section, a literature review is provided which discusses different approaches for AQI forecasting. The third section is devoted to the presentation of GP and four machine learnings used for

prediction. The fourth section describes the data, the pre-possessing steps, and the metrics to evaluate models. In section 5, the performances of different approaches are reported, and the results are compared. Finally, Section 6 is dedicated to the conclusions.

2. LITERATURE REVIEW

A regression model was used to estimate PM_{10} concentration in Chonburi, Thailand, using meteorological and pollutant data. The model included air pressure, precipitation, temperature, relative humidity, and wind speed. The pollutants included carbon monoxide, nitrogen monoxide, nitrogen dioxide, sulfur dioxide, Black Carbon, methane, Non-Methane Hydro Carbon, and ozone. Another study by Rybarczyk and Zalakeviciute [7] estimated $PM_{2.5}$ concentration using regression models based on time in Quito, Ecuador. The model showed an R^2 score of 0.27 with these settings. By adding meteorological data, the score improved to 0.38. The study also considered trace gas concentrations, improving the R^2 score to 0.8. The limitation was the extra cost of measuring trace gas concentrations.

Support vector regression was used to estimate CO concentration in New South Wales, Australia, using four different sets of features including CO concentrations from four monitoring stations, latitude and longitude, hour, day of the week, and season [8]. The same spatial pollutant monitoring networks were employed in modelling. Support vector regression was also applied to predict O_3 concentration in Delhi, India, using different kernels such as linear, polynomial, and radial basis functions. The best feature set for forecasting O_3 concentration was found with five input parameters: ozone for the previous two days and meteorological inputs of air temperature, relative humidity, and sunshine hours. A comparison between linear regression and multiple layer perceptron performance metrics concluded that support vector regression effectively captured non-linear trends. Similar experiments were conducted to forecast the air quality index in Tehran from 2008 to 2013, exploring different kernel functions and obtaining a pollutant map with different AQIs for different locations [9].

Ensemble methods also are widely used in air pollution estimation due to their popularity and applicability. These methods consider meteorological parameters such as air temperature, air pressure, relative humidity, and wind speed as input parameters. These variables vary depending on location and play a crucial role in rapidly varying pollutant concentrations. In Delhi, India, eleven models were used for $PM_{2.5}$ forecasting, with the outputs from two different models combined to improve performance [10]. Another study compared 23 features and $PM_{2.5}$ concentration from 37 monitoring stations, finding that the performance capabilities of artificial neural networks were reduced due to missing values in $PM_{2.5}$ and aerosol optical depth data [11].

A random forest model for ozone estimation was built at the Research Academy for Environmental Sciences in Beijing, China [12]. A linear hybrid machine learning model was applied for $PM_{2.5}$ concentration estimation in China [13]. In London, $PM_{2.5}$ was estimated using PM_{10} and NO_X emissions using regression modeling, machine learning methods, and a combination of both [14].

Six machine learning models were investigated to estimate the prediction capability of $PM_{2.5}$ and NO_2 [15]. Daily CO concentration was estimated in Taiwan from 2000 to

2018, with three models using a deep neural network, random forest, and XGBoost. The XGBoost model had the highest R^2 score of 0.85, followed by random forest and neural network with 0.84 and 0.81 respectively [16]. Satellite-Based estimates of daily NO_2 exposure in China were tested using a hybrid random forest and spatiotemporal kriging model [17].

Isam Drewil and Jabbar Al-Bahadili's 2022 study proposed a model that combines the Genetic Algorithm (GA) with Long Short Term Memory (LSTM) to optimize hyperparameters and predict pollution levels for the next day. The metaheuristic GA offers a flexible solution to the primary challenge of selecting appropriate parameters for LSTM. Du et al. [18] reviewed the effectiveness of four advanced machine learning methods for spatial data handling, including SVM, semi-supervised and active learning, ensemble learning, and deep learning. Filonchyk et al. [19] investigated spatial and temporal variations of atmospheric pollutants using satellite-based measurement data and ground-based results. Kumar et al. [20] evaluated different interpolation techniques for air quality mapping in Mumbai, India, and found that the IDW method performed better in statistical assessments, suggesting that the IDW approach performs favorably among the interpolation techniques tested in this study.

3. METHODOLOGY

In this paper, the proposed methods use GP to draw a comparative analysis of the AQI values.

3.1. Genetic Programming

GP is a heuristic search approach based on program evolution, displaying programs as tree structures that learn and adapt. It uses symbolic regression to search for relationships between variables and evolves functions until a solution is reached. GP is similar to genetic algorithms but uses computer programs or equations. Fundamental procedures like reproduction, mutation, and crossover are explained in Koza [21].

Examples of genetic operations

- Generating population

A program is given in Fig. 1 in the form of a tree structure. The programs are represented by a population of random trees, and genetic operations are conducted on these trees to create individuals using two separate sets: the terminal set T and the function set F. {*, /, -, √} ⊆ F and {x, y, z, 7.2} ⊆ T for Fig. 1. To construct a random tree, choose at random from T∪F until all branches end up in terminals.

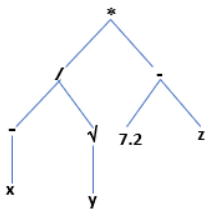

Figure 1. *Program in the form of a tree representation.*

- Crossover

Two random nodes are chosen from inside such a program (parents), and the resulting sub-trees are swapped, resulting in two new programs.

- Mutation

A sub-tree is replaced by another one randomly.

- Reproduction

This means an exact duplication of the program if it is found to be acceptable by the fitness criteria.

GP generates a random population of individuals, assesses their fitness, and chooses 'parents'. They produce offspring through reproduction, mutation, and crossover. The process continues iteratively until a certain number of offspring are produced, and the solution is the outcome of this process. The step-by-step approach for implementing the GP is further outlined below:

1. Generate an initial random population of individuals (equations or programs) of a certain size by selecting them at random from a set of terminals and functions.
2. Evaluate the fitness of each individual in a population using a criterion such as the RMSE.
3. Choose individuals or parents (often through a tournament including comparing two parents at a time and then shortlisting the winner for further competition).
4. Create additional offspring (individuals) from these parents by doing the operators: Replication, Crossover or Mutation.
5. If the number of individuals (offspring) equals the maximum specified number (or found the optimal solution), raise the number of generations by one and go to step 6; otherwise, repeat steps 2-5 to increase the number of individuals.
6. If the number of generations reaches a specific maximum, the program is terminated; otherwise, repeat steps 2-5.

3.2. Other machine learning methods

Support vector machine - a supervised learning model for classification and regression, artificial neural network-learning methodology inspired by actual neurons of the brain, random forest -techniques utilizing an ensemble of weak prediction models, and k-nearest neighbor-a lazy learning nonparametric supervised method are some of the existing algorithms used to predict AQI in this paper.

4. EXPERIMENTAL SETUP

In this section, we discuss the implementation of the GP and ML methods for comparison.

4.1. Dataset

The dataset utilized for this experiment is linked https://www.kaggle.com/rohanrao/air-quality-data-in-india.

The dataset contains hourly and daily air quality and AQI data from a variety of stations in various Indian cities. The statistics cover the years 2015 through 2020. The original dataset has 29532 rows and 16 columns, containing all of the cities.

The attribute data is shown: City, $PM_{2.5}$, PM_{10}, NO, NO_2, NO_x, NH_3, CO, SO_2, O_3, Benzene, Toluene, AQI, and AQI_Bucket (however, the AQI bucket is not the value we want to predict in this study so we do not use this value).

The dataset is cleaned and picked from the original dataset's 6 stations of city datasets: Amritsar, New Delhi, Hyderabad, Amaravati, Kolkata, Visakhapatnam and Hyderabad as shown in table 1.

The dataset will be prepared, cleaned, reduced, normalized, and divided into training and testing data. The goal is to use simple implementations for real-world use. Two metrics will be used to evaluate and compare algorithms, determining accuracy and selecting the best one.

4.2. Statistical evaluation of model performance

The predicted and observed AQI values were compared using two measures of predict accuracy computed from the test datasets: the Normalized Root Mean Squared Error (NRMSE), and the Coefficient Correlation (also called *r*).

Table 1. *The stations selected from the India data set.*

No.	Station code	Station name	City	State
1.	PB001	Golden Temple, Amritsar - PPCB	Amritsar	Punjab
2.	TG001	Bollaram Industrial Area, Hyderabad - TSPCB	Hyderabad	Telangana
3.	WB009	Fort William, Kolkata - WBPCB	Kolkata	West Bengal
4.	DL001	Alipur, Delhi - DPCC	Delhi	Delhi
5.	AP005	GVM Corporation, Visakhapatnam - APPCB	Visakhapatnam	Andhra Pradesh
6.	AP001	Secretariat, Amaravati - APPCB	Amaravati	Andhra Pradesh

Table 2. *Datasets of the stations.*

No.	Station code	Dataset size	Training set	Testing size
1.	PB001	642	428	214
2.	TG001	907	605	302
3.	WB009	384	256	128
4.	DL001	467	311	156
5.	AP005	1145	763	382
6.	AP001	630	420	210

The NRMSE is defined as:

$$RMSE = \frac{\sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2}}{y_{max} - y_{min}}$$

The r is defined as

$$r = \frac{\sum_{i=1}^{m}(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^{m}(y_i - \bar{y})^2 \sum_{i=1}^{m}(\hat{y}_i - \bar{\hat{y}})^2}}$$

where $y_i,\ i = 1 \div m$ are observed values for the AQI, $\bar{y}$ is their mean and $\hat{y}_i,\ i = 1 \div m$ are predict values, $\bar{\hat{y}}$ is their mean.

The sign of r indicates the strength of the linear relationship between variables, with a strong connection near 1, zero, or -1 indicating no relationship.

5. RESULTS AND ANALYSIS

In this section, the paper will conduct training and validate the correctness of building a GP prediction model. For each data set corresponding to the city station, we built 5 separated models for each algorithm and then used NRMSE calculates the deviation between observed data and predicted data, while CC is the correlation coefficient (r) between actual and predicted numbers.

Table 3. *GP parameter settings.*

Parameter	**Value**
Function set	+, -, *, / (protected division), sin, cos, exp, lg, √
Variable terminals at	all features
Constant terminals	Random double values range from -500 to 500
Population size	1000
Initialization	Ramped half-and-half
Generations	200
Crossover probability	60%
Mutation probability	30%
Reproduction rate	10%
Selection type	Tournament (size=3)

Table 4. *Optimal parameters using Weka for the four models: SVM, k-NN, MLP and RF for AQI forecasting.*

SVM	**k-NN**
SVM Type: epsilon-SVR Kernel Type: RBFKernel cacheSize: 250007 gamma: 0.01 Epsilon:1.0E-12	k:5 distanceFunction: Euclidean
MLP	**RF**
hiddenLayers: 5 learningRate: 0.3 Momentum: 0.2 epochs: 500	Max depth 0 (unlimited) # features int(log2(predictors) + 1) # iterations 100

As mentioned, we will compare the results of GP with other methods such as Support Vector Machines, k Nearest Neighborhood, Multilayer Perceptron, REF, or more specifically SMOreg, Ibk, MLP and RF in Weka [16]. GP will be implemented by using ECJ [22]. Specific parameters for each method are as follows in table 3 and table 4.

5.1. Performance comparison of algorithms

In this subsection, the performance of GP and four ML techniques is examined using two different evaluation metrics. It can be seen from the table 5, 6 that the NRMSE and r values of GP are all better than the other models. If we use NRMSE as the evaluation criterion, the GP model has the best performance, followed by MLP, RF, SVM and kNN. Similarly, the r value of GP is larger than that of MLP, RF, SVM and kNN. Fig. 2 to Fig. 7 show the graphs of the prediction of GP and ML models and the for 6 datasets.

The larger part of the models are scattered on the CC values from 85% to 91% and the NRMSE values from 9% to 15% in Fig. 2, all models in Fig. 3 have CC values larger than 86%. Similarly, in Fig. 3 all models have CC values larger than 90%. In Fig. 4 and Fig. 5, only the CC value of the kNN model has not exceeded the value of 80%, while the other models have their CCs all greater than 83%. The last figure shows that the CC values of the models are very large, almost over 90%. Among them, the GP appeared the most excellent performance both in the values of CC and NRMSE.

By comparing the experimental results of the GP and four models, the GP method is more efficient for AQI prediction so it is a good alternative to existing models for AQI forecasting.

5.2. The evolved model of GP

The best model evolved by GP is shown following for WB009 station:

$$sin\sqrt{sinx_1 * x_2} - 431.02 * 334.93log\sqrt{x_9}$$

Within x_1, x_2, x_9 respectively stand for the $PM_{2.5}$, PM_{10}, O_3. The above equation implies that GP can automatically pick features to evolve the best models. This clearly highlights the significance of the GP white-box.

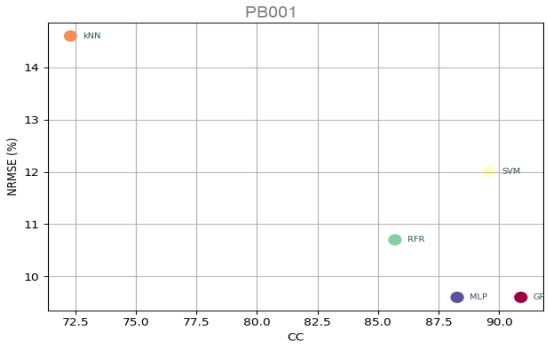

Figure 2. *Diagram of the correlation coefficient (CC) and the NRMSE (%) for the AQI forecast made in PB001 station by 4 models and GP.*

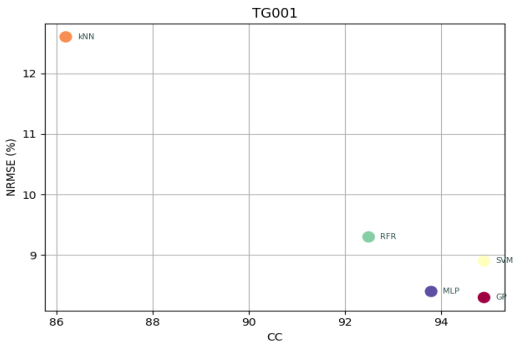

Figure 3. *Diagram of the correlation coefficient (CC) and the NRMSE (%) for the AQI forecast made in TG001 station by 4 models and GP.*

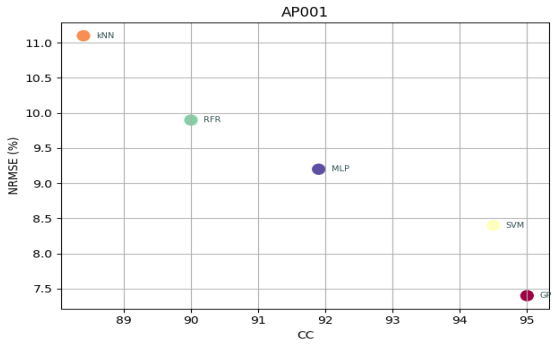

Figure 4. *Diagram of the correlation coefficient (CC) and the NRMSE (%) for the AQI forecast made in AP001 station by 4 models and GP.*

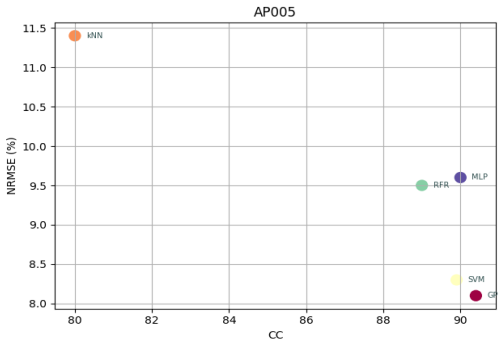

Figure 5. *Diagram of the correlation coefficient (CC) and the NRMSE (%) for the AQI forecast made in PB001 station by 4 models and GP.*

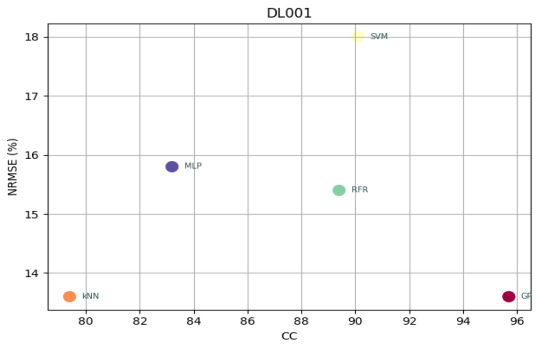

Figure 6. *Diagram of the correlation coefficient (CC) and the NRMSE (%) for the AQI forecast made in PB001 station by 4 models and GP.*

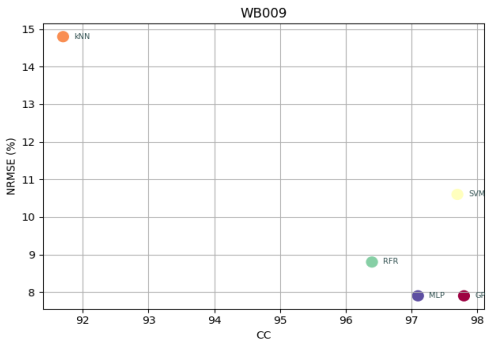

Figure 7. *Diagram of the correlation coefficient (CC) and the NRMSE (%) for the AQI forecast made in PB001 station by 4 models and GP.*

6. CONCLUSION AND FUTURE WORKS

Air pollution is a worldwide issue, and experts from all around the world are attempting to find a solution. Machine learning approaches were researched in order to properly anticipate the AQI. The applicability of the Genetic Programming approach to address the AQI forecasting problem may be proven based on the following experimental comparison findings. Not only does the approach exceed existing techniques in terms of accuracy, but also the GP method suggested in the research is fully data independent, employs a white-box strategy, and can be applied to forecasting in a wide range of locales.

The derived results of this research can be applied in practice to forecast future AQI of any day with a similar pattern. Moreover, by performing the necessary data pre-processing and tuning the hyperparameters, the provided models can be applied to many different datasets. This has significant implications for advanced strategic decision-making and planning for environment.

There are plans for future work to employ satellite images and more thorough data to offer estimates for particular regions of a city. Another way to investigate is artificial intelligence (AI) to improve the models' effectiveness and innovation.

Table 5. *The NRMSE value of GP and 4 models (%).*

No.	Station code	GP	MLP	SVM	kNN	RFR
1.	**PB001**	**9.6**	**9.6**	12	**14.6**	**10.7**
2.	**TG001**	**8.3**	**8.4**	8.9	12.6	9.3
3.	**WB009**	**7.9**	**7.9**	10.6	14.8	8.8
4.	**DL001**	**13.6**	15.8	18	13.6	15.4
5.	**AP005**	**8.1**	9.6	**8.3**	11.4	9.5
6.	**AP001**	**7.4**	9.2	8.4	11.1	9.9

Table 6. *The* $\boldsymbol{r}$ *value of GP and 4 models (%).*

No.	Station code	GP	MLP	SVM	kNN	RFR
1.	**PB001**	**90.9**	88.27	89.63	72.28	85.74
2.	**TG001**	**94.9**	93.8	94.9	86.17	92.49
3.	**WB009**	**97.8**	97.1	97.7	91.7	96.4
4.	**DL001**	**95.7**	83.2	90.1	79.4	89.4
5.	**AP005**	**90.4**	90	89.9	80.0	89.0
6.	**AP001**	**95.0**	91.9	94.5	88.4	90.0

REFERENCES

[1]. E. E. A. (EEA), *"Air Quality in Europe 2022 Report,"* Publications Office, https://doi.org/10.2800/488115. ISBN: 978-92-9480-515-7, (2022).

[2]. J. v. d. H. D. Z. P. v. R. S. Duyzer, *"Representativeness of air quality monitoring networks,"* Atmos. Environ., vol. 104, p. 88–101, (2015).

[3]. G. M. A. M. W. P. E. &. A. E. Raimondo, *"A machine learning tool to forecast PM10 level,"* in AMS 87th Annual Meeting, San Antonio, TX, USA, (2007).

[4]. G. Raimondo, A. Montuori, W. Moniaci, E. Pasero and E. Almkvist, *"A Machine Learning Tool to Forecast PM10 Level,"* in The Fifth Conference on Artificial Intelligence Applications to Environmental Science, San Antonio, TX, USA, (2007).

[5]. R. Y. Y. Y. L. H. G. &. M. O. A. Yu, *"RAQ–A random forest approach for predicting air quality in urban sensing systems,"* Sensors, vol. 16, no. 86, (2016).

[6]. K. &. D. A. Veljanovska, *"Air quality index prediction using simple machine learning algorithms,"* International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), vol. 7, no. 1, pp. 25-30, (2018).

[7]. H. Q. L. D. Y. a. Y. G. Liu, *"Air quality index and air pollutant concentration prediction based on machine learning algorithms,"* Applied Sciences, vol. 9, p. 4069, (2019).

[8]. K. S. V. B. H. K. S. R. A. Hu, *"Svr based dense air pollution estimation model using static and wireless sensor network,"* IEEE SENSORS, pp. 1-3, (2016).

[9]. A. Chelani, *"Prediction of daily maximum ground ozone concentration using support vector machine,"* Environmental monitoring and assessment, vol. 162, pp. 169-176, 2009.

[10]. S. M. S. S. S. Kumar, *"A machine learning-based model to estimate pm2.5 concentration levels in Delhi's atmosphere,"* Heliyon 6, vol. 11, (2020).

[11]. M. C. C. N. X. B. B. T. S. Zamani Joharestani, *"PM2.5 prediction based on random forest, xgboost, and deep learning using multisource remote sensing data,"* Atmosphere, vol. 10, no. 7, p. 373, (2019).

[12]. J. L. Y. M. W. Z. X. W. X. B. F. Z. Y. W. Z. L. H. Zhan, "Ozone formation sensitivity study using machine learning coupled with the reactivity of volatile organic compound species," Atmospheric Measurement Techniques, vol. 15, no. 5, pp. 1511-1530, (2022).

[13]. Z. C. B. H. Y. D. L. Y. T. Song, *"Estimation of pm2.5 concentration in China using linear hybrid machine learning model,"* Atmospheric Measurement Techniques, vol. 14, no. 8, p. 5333–5347, (2021).

[14]. A. B. B. G. D. B. A. S. E. S. J. K. K. Analitis, *"Prediction of pm2.5 concentrations at the locations of monitoring sites measuring pm10 and nox, using generalized additive models and machine learning methods: a case study in london,"* Atmospheric Environment, vol. 240, (2020).

[15]. Z. Y. S.-L. H. K.-F. Li, *"High temporal resolution prediction of street-level pm2.5 and nox concentrations using machine learning approach,"* Journal of Cleaner Production, vol. 268, (2020).

[16]. P.-Y. H. C.-Y. W. J.-Y. T. T.-A. H. J.-W. G. H.-R. S. H.-J. W. C.-. D. S. J. Wong, *"Incorporating land-use regression into machine learning algorithms in estimating the spatial-temporal variation of carbon monoxide in taiwan,"* Environmental Modelling Software, vol. 139, (2021).

[17]. Y. L. Y. D. X. Z. K. Z. M. G. M. d. B. Zhan, *"Satellitebased estimates of daily no2 exposure in China using hybrid random forest and spatiotemporal kriging model,"* Environmental Science Technology, vol. 52, no. 3, (2019).

[18]. P. B. X. T. K. X. Z. S. A. X. J. L. E. S. H. L. W. Du, *"Advances of four machine learning methods for spatial data handling: a review,"* Journal of Geovisualization and Spatial Analysis, vol. 4, pp. 1-25, (2020).

[19]. M. H. V. Y. H. &. Y. S. Filonchyk, *"Atmospheric pollution assessment near potential source of natural aerosols in the South Gobi Desert region, China,"* GIScience & Remote Sensing, vol. 57, no. 2, pp. 227-244, (2020).

[20]. A. D. S. &. D. A. K. Kumar, *"Comparative evaluation of fitness of interpolation techniques of ArcGIS using leave-one-out scheme for air quality mapping,"* Journal of Geovisualization and Spatial Analysis, vol. 6, no. 1, p. 9, (2022).

[21]. J. Koza, Genetic Programming: On the Programming of Computers by Means of Natural Selection, MIT Press, (1992).

[22]. L. P. G. B. S. P. Z. S. J. B. R. H. a. A. C. S. Luke, *"Ecj: A java-based evolutionary computation research system,"* (2007).

[23]. M. H. V. Y. H. &. Y. S. Filonchyk, *"Atmospheric pollution assessment near potential source of natural aerosols in the South Gobi Desert region, China,"* GIScience & Remote Sensing, vol. 57, no. 2, pp. 227-244, (2020).

TÓM TẮT

Dự báo chỉ số chất lượng không khí dựa trên lập trình di truyền

Chỉ số chất lượng không khí (AQI) là công cụ dùng để đo lường tác động của ô nhiễm không khí đối với sức khỏe con người theo thời gian. Ô nhiễm không khí đang gia tăng đáng kể trên thế giới và gần đây các kỹ thuật học máy được sử dụng để dự báo và phân tích AQI. Chúng tôi trình bày một cách tiếp cận mới, sử dụng GP để phát triển các mô hình dự báo AQI. GP có thể phát triển các mô hình dự báo AQI chính xác hơn một số thuật toán học máy phổ biến, theo kết quả thử nghiệm sử dụng bộ dữ liệu thu được từ nhiều trạm khác nhau trên nhiều thành phố ở Ấn Độ. Hơn nữa, trong khi phát triển các mô hình dự báo AQI, GP có thể tự động xác định các đặc trưng quan trọng và chính vì vậy các mô hình tiến hoá bởi GP là có tính giải thích.

Từ khóa: Học máy; Dự báo AQI; GP; Lập trình di truyền.