IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 5 June 2023, accepted 21 June 2023, date of publication 30 June 2023, date of current version 7 July 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3291146

RESEARCH ARTICLE

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

CHUNHAO LIU, (Graduate Student Member, IEEE), GUANGYUAN PAN, (Member, IEEE), DONGMING SONG, AND HAO WEI
School of Automation and Electrical Engineering, Linyi University, Linyi 276000, China

Corresponding author: Dongming Song (david.dming@qq.com)

The research conducted in this study was supported by National Natural Science Foundation of China under Grant no. 62103177. This funding played a crucial role in facilitating data collection, conducting experiments, and analyzing the results.

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 (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.

INDEX TERMS Time series, air quality forecasting, machine learning, extreme learning machine, genetic algorithm.

I. 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 [1]. Li et al. came to the conclusion that outdoor physical activity poses numerous health risks due to ambient air pollution in China. [2], [3]. According to the Chinese Ambient Air Quality Standards (GB3095-2012), there are six conventional air pollutants used to measure air quality: sulfur dioxide (SO_2), nitrogen dioxide (NO_2), particulate matter with a particle size less than 10 microns (PM_{10}), particulate matter with a particle size less than 2.5 microns ($PM_{2.5}$), ozone (O_3), and carbon monoxide (CO) [4], [5], [6]. 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.,$PM_{2.5}$) or traffic-related pollutants, is linked to higher rates of lung cancer, coronary heart disease, and

The associate editor coordinating the review of this manuscript and approving it for publication was Frederico Guimarães.

other illnesses [7], [8]. 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.

Recently, the inherent property of machine learning algorithms to automatically learn features at multiple levels of abstraction has become increasingly important in providing solutions to this challenging task [9], [10]. However, the model only forecasts PM_{10} and SO_2 levels, and it is also challenging to obtain measurement values needed to construct the dataset [11]. Wu Q. et al. proposed an optimal-hybrid model for daily AQI prediction considering air pollutant factors, with the model's inputs being the six atmospheric pollutants. However, neural networks typically struggle with slow learning, a tendency to fall into local minima, and a complex network training process. Based on the generalized inverse matrix theory, Huang et al. proposed an extreme learning machine (ELM) algorithm with a feedforward neural network that includes a single hidden layer, such that the problems of conventional neural network algorithms are circumvented. The ELM algorithm used to predict the AQI outperformed neural networks in terms of parameter selection, training speed, and prediction accuracy [12]. However, the parameters of the hidden layer nodes and the number of nodes in the test hidden layer are selected at random, which puts the prediction accuracy to a great test.

In order to solve the aforementioned problems, we propose to optimize the number of ELM hidden layer nodes, thresholds, and weights, along with an improved genetic algorithm (GA) that uses root mean square error (RMSE) as the fitness function, to obtain the optimal network structure for air quality prediction [14]. The number of hidden layer nodes is updated by continuous coding discretization, the input weights and hidden layer thresholds are updated by continuous coding, and the update thresholds and weights are selected with the number of updated layers to form a hierarchical control structure [15]. The proposed GA-based improved extreme learning machine (GA-KELM) algorithm is applied to air quality prediction, and its performance is compared with that of community multiscale air quality modeling system (CMAQ), support vector regression (SVR), and deep belief network-back propagation (DBN-BP). The results show that the accuracy of the proposed GA-KELM algorithm is reliable for air quality prediction [16].

In this study, an improved extreme learning machine model based on a genetic algorithm is designed and applied to AQI prediction. To verify the effectiveness of the model, we conducted tests on three real-world datasets. The results confirmed that the proposed method has superior performance and outperforms some advanced methods currently in use. The main contributions of this paper are: (1) modifying the ELM activation function or using the kernel function to improve the prediction accuracy, (2) optimizing the ELM using GA to improve the stability of the results and further enhance the prediction accuracy, and (3) obtaining the correlation analysis results of atmospheric environmental quality prediction parameters by comprehensively considering each relevant factor in line with the actual situation. The remainder of this paper is organized as follows. Section II presents related work. Section III describes ELM and the proposed GA-KELM, and illustrates the improvements using the model. Section IV discusses experimental results where GA-KELM is compared with several other methods in terms of prediction results. The last section concludes the entire work and presents directions for future research.

II. RELATED WORK

Air quality prediction has been extensively researched in the literature [17]. In recent years, numerous researchers have made significant contributions to the field by leveraging quantitative studies and the latest techniques to identify various air quality patterns and their underlying trends [18]. Existing work in this area relies on statistical methods and shallow machine learning models to address the problem of air quality prediction [19].

Agarwal and Sahu [20] conducted air quality prediction studies by employing statistical models. Lary et al. [21] combined remote sensing and meteorological data with ground-based $PM_{2.5}$ observations. Zheng et al. [22] proposed a hybrid prediction method that combines a linear regression- based temporal prediction method with an ANN-based spatial prediction method for pollutant concentrations. Zheng et al. [23] used a data-based approach for the next 48 hours of $PM_{2.5}$ prediction, implementing a prediction model based on linear regression and neural network. They combined meteorological data, weather forecast data, and air quality data from monitoring stations. Rajput and Sharma [24] used a multiple regression model to represent the changes in air quality index (AQI), considering ambient temperature, relative humidity, and barometric pressure as the main parameters in the regression model for AQI calculation [25]. These classical methods and models all have the advantages of simple algorithms, easy processing, and acceptable prediction results. However, obtaining precise and specific air quality prediction values remains challenging [26].

Elbaz et. al. [27] proposed a novel deep learning approach that extracts high-level abstractions to capture the spatiotemporal characteristics of NEOM city in Saudi Arabia at hourly and daily intervals. Campbell et al. [28] described the development of FV3GFSv16 coupled with the ''state-of-the-art'' CMAQ model version 5.3.1. Jin et al. [29] proposed an interpretable variational Bayesian deep learning model with self-filtering capability for $PM_{2.5}$ prediction information, which effectively improves prediction accuracy.

Zhou et al. [30], [31], [32] proposed a method based on an improved Grasshopper optimization algorithm to classify the color difference of dyed fabrics using kernel extreme learning machine. In this study, the classification of color differences in dyed fabric images is performed using the kernel limit learning machine, and the kernel function parameters are optimized by the improved Grasshopper optimization algorithm to achieve color difference classification of dyed fabric images. Xue et al. [33] proposed a GA-based air quality prediction model to optimize the parameters of the weighted extreme learning machine (WELM). Despite the progress made by the aforementioned methods, they also exhibit limitations; their training efficiency is relatively low, and deep learning algorithms are not yet fully mature. These challenges present greater obstacles for the application of deep learning, necessitating improvements to existing models, the development of new models, and the enhancement of their predictive capabilities [34], [35].

The use of statistical or numerical forecasting techniques is subject to several limitations. Neural networks are widely used because of their unique associative abilities, memory, and distinctive learning [36], [37]. Given the highly nonlinear nature of AQI changes and the strong generalization and nonlinear characterization abilities of neural networks, the nuclear limit learning machine neural network model, also known as kernel extreme learning machine (KELM), is employed to investigate air quality prediction using a real dataset. The weights and threshold values of KELM are optimized using a genetic optimization algorithm [38].

III. METHODOLOGY

In this section, 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 [39]. 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.

The AQI of each pollutant, or individual AQI (IAQI), is the largest value of the day. AQI is calculated as in (1):

$$AQI = \max\{LAQI_1, L\ AQI_2, L\ AQI_3, \cdots, L\ AQI_n\} \quad (1)$$

where $LAQI_1, L\ AQI_2, L\ AQI_3, \cdots, L\ AQI_n$ represent the sub-index corresponding to each pollutant item. In this study, the calculation of AQI involves only six pollutants.Therefore, (1) can be expressed as:

$$AQI = \max\{IAQI_{SO_2}, IAQI_{N_2}, IAQI_{O_3}, IAQI_{PM_{10}}, IAQI_{PM_{2.5}}, IAQI_{CO}\} \quad (2)$$

To meet our research goals, this paper proposes GA-KELM, a method for air quality prediction based on an improved extreme learning machine, which in turn is based on an improved GA [40]. Aiming at the problem of network instability caused by the randomly generated input layer weights and hidden layer thresholds of KELM, a GA is used to optimize the KELM weights and thresholds, thereby improving the model's performance in terms of prediction accuracy, which is the main objective of this algorithm. In each iteration of the GA, a new offspring population is generated by selection, crossover, and mutation, and the individual with good fitness value is selected. The GA stops iterating when the stopping criteria are satisfied. The GA is used to determine the optimal weights and threshold values, which overcomes the instability of KELM and reduces prediction errors, thus resulting in a more reliable prediction model and improved air quality prediction accuracy. Details of the model will be discussed in the following sections.

A. EXTREME LEARNING MACHINE

ELM was first proposed by Huang. It is characterized by its fast training and high training accuracy. Feedforward neural networks are mainly based on the gradient descent method [41]. Their main drawbacks are the slow training, the tendency to fall into a local minimum point that cannot reach the global optimum, and the high sensitivity to the learning rate η (if the selected rate is improper, it might cause slow convergence and the training process will thus take a long time, or it becomes unstable). FIGURE 2 shows the network structure of an ELM.

Consider N different samples $(x_i, t_i), i \in 1, 2, \cdots N$, where x_i denotes the input and t_i represents the target, L hidden layer neurons, and an activation function $g(x)$.The mathematical expression for ELM output is:

$$y_i = \sum \boldsymbol{\beta}_j g\left(w_j x_i + b_j\right) \quad (3)$$

where $j \in 1, 2, \cdots, L$, w_j are the weights of the input and hidden layer neurons,b_j are the thresholds of the hidden layer neurons, β_j is the weight matrix of neurons in the hidden and output layers, and $g\left(w_j x_i + b_j\right)$ is the output of the hidden layer neurons.

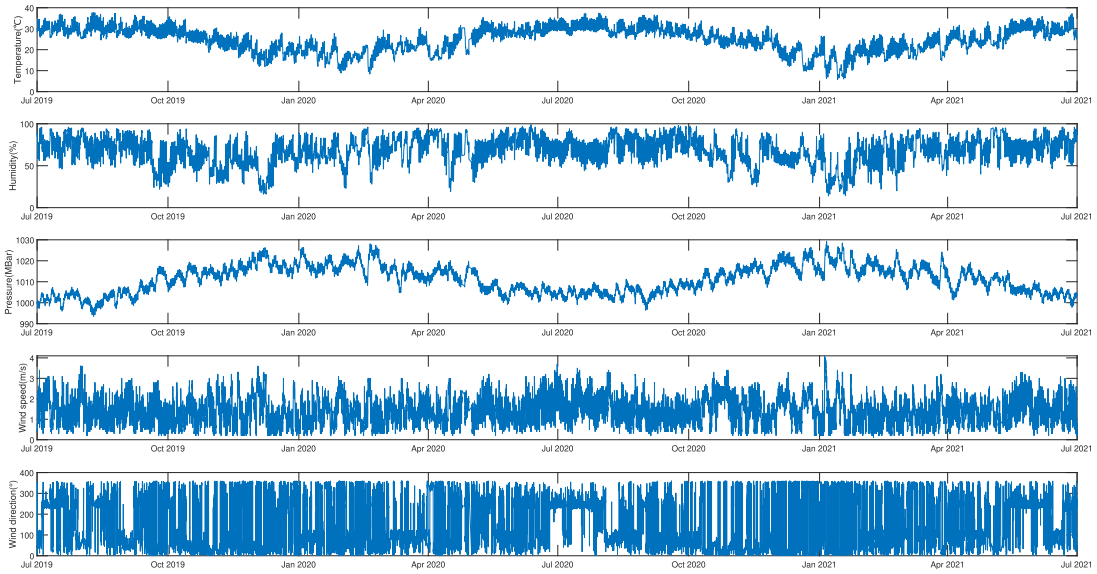

FIGURE 1. Sequence diagram of meteorological factors.

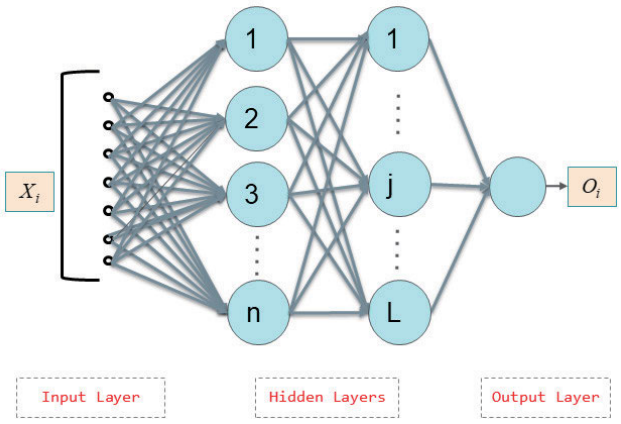

FIGURE 2. Schematic diagram of ELM structure.

In order to approximate the ideal output result, there must be a set of w_j, b_J and β_j such that:

$$\sum_{i=1}^{N} \boldsymbol{\beta}_j g\left(w_j x_i + b_j\right) = t_i, i = 1, 2, \cdots, N \tag{4}$$

Like the single hidden layer neural network, (4) can be expressed as:

$$H\beta = T \tag{5}$$

where H is the output matrix of the hidden layer neurons and T is the target output of the ELM network, i.e.,$\mathbf{T} = \{t_i\}_{i=1}^{N}$, so (5) can be transformed into the least squares problem of solving for the weight matrix, namely:

$$\|\mathbf{H}\hat{\boldsymbol{\beta}} - \mathbf{T}\| = \min_{\beta} \|\mathbf{H}\boldsymbol{\beta} - \mid \mathbf{T}\| \tag{6}$$

The output weight matrix $\hat{\boldsymbol{\beta}}$ can be obtained as:

$$\hat{\boldsymbol{\beta}} = H^{+}T \tag{7}$$

where H^{+} is the Moore-Penrose generalized inverse of H. When the number of training observations is greater than the number of hidden nodes, the expression β of the calculated output weight is given by:

$$\hat{\beta} = \left(\frac{I}{C} + H^T H\right)^{-1} H^T T \tag{8}$$

where I is the K-dimensional identity matrix and $\frac{I}{C}$ is the regularization parameter. When the number of training samples is less than the number of hidden nodes, the output weight $\hat{\beta}$ expression is written as:

$$\hat{\beta} = H^T \left(\frac{I}{C} + HH^T\right)^{-1} T \tag{9}$$

B. IMPROVED ELM BASED ON KERNEL METHOD

Inspired by SVM, the kernel method is introduced into ELM, namely KELM. KELM maps linearly inseparable patterns to a high-dimensional feature space to achieve linear separability and improve the accuracy and robustness of the model [42]. Therefore, it ensures that the network produces good generalization and learns fast, and it avoids the shortcomings of gradient descent training algorithms that are prone to falling into local optima and necessitate a large number of iterations.

ELM is a single hidden layer feedforward neural network, and its learning objective (10) can be expressed as a matrix:

$$F(x) = h(x)\beta = H\beta = L \tag{10}$$

Turning network training into a problem of solving a linear system, β is determined by $\beta = H * L$,where H^* is a generalized matrix of H.In order to enhance the stability of the model, the regularization coefficient C and the identity matrix I are introduced, and the output weights can be

calculated by the least square's solution in (11).

$$\beta = H^T \left(HH^T + \frac{I}{c} \right)^{-1} L \tag{11}$$

In addition to the many advantages of the ELM algorithm, KELM combines the kernel function to achieve linear separability by non-linearly mapping the linearly inseparable pattern to the high-dimensional feature space, which further improves the judgmental accuracy. The kernel function is introduced into ELM, and the kernel matrix is:

$$\Omega_{ELM} = HH^T = h(x_i)\, h(x_j) = K(x_i, x_j) \tag{12}$$

where x_i and x_j represent test input vector. Therefore,(10) can be expressed as:

$$F(x) = [K(x, x_1); \ldots; K(x, x_n)] \left(\frac{I}{C} + \Omega_{ELM} \right)^{-1} L \tag{13}$$

where $x_1, x_2, \cdots, x_n$is a given training sample, n is the number of samples, and $K(x_1, x_i)$ is the kernel function, $i = 1, 2, \cdots, n$. One of the widely used kernel functions is the radial basis function (RBF). The RBF kernel can be defined as:

$$K(x, y) = \exp\left(-\gamma \|x - y\|^2 \right) \tag{14}$$

where y is the kernel parameter. Since the results of the KELM model strictly depend on the regularization parameter C and the choice of the kernel parameter y, these two parameters need to be optimized effectively.

In conventional ELM, input weights and hidden layer neurons are assigned at random, and output weights are then obtained on the basis of input weights and hidden layer neurons. Thus, human experiments are essentially important in customary training, which often results in inaccurate predictions [43]. To overcome this shortcoming, this paper considers using a GA to optimize input weights and model structure. In this section, as the input weights are continuous variables and the model structure is discrete, the GA is improved to handle the difference between continuous and discrete variables of KELM in parallel optimization. The improved GA-based KELM process is shown in FIGURE 3.

The steps for using a GA to optimize the parameters of KELM are as follows:

Step 1: Data normalization. Since the observed data has different properties, dimensions, and obvious differences in numerical values, normalization is performed so that the distribution range of the processed data is kept between [0, 1], in order to avoid learning errors caused by the bias. The normalization formula is given by:

$$X = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{15}$$

where x is the observed value, X is the input data of the prediction model,x_{min} and x_{max} are the minimum and maximum of the observed values, respectively [44].

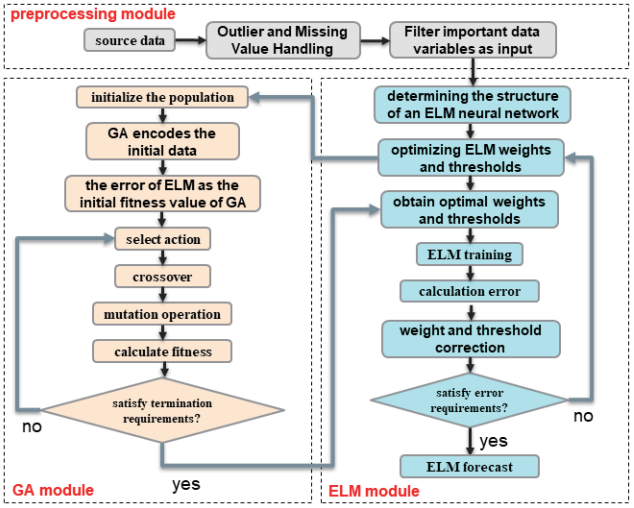

FIGURE 3. GA-KELM algorithm flowchart.

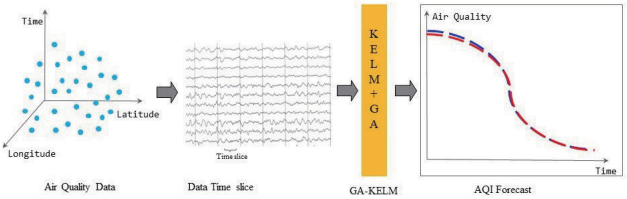

FIGURE 4. Model architecture diagram.

Step 2: The input weights w and thresholds b of the hidden layer of the ELM network are randomly generated as the initial population of GA.

Step 3: The KELM model is trained using the training samples, and the trained network is used to predict test samples. The testing result is sent to the GA as objective function.

Step 4: The improved GA is used to optimize weights and model structure.

Step 5: Steps 3-4 are repeated until convergence is reached.

Step 6: The KELM is assigned the optimal hidden layer input weights w and thresholds b for the final training and testing.

C. OVERVIEW OF THE DEEP AIR QUALITY FORECASTING FRAMEWORK

The flowchart of applying GA-KELM for AQI prediction is shown in FIGURE 4.

a) Data collection and preparation. In GA-KELM, before performing GA, the fitness value obtained during the prediction process of the initial KELM is considered.

b) Prediction. In the experiment, two-thirds of the collected data are used to train the KELM network, and the remaining one-third is used as a test set.

c) Optimization. The GA is applied through the processes of selection, crossover and mutation, until the optimal result is obtained. The weights and structure parameters of KELM are then updated using the values produced by the GA. This process is repeated until the optimal result is achieved.

All the spatiotemporal relationships that dominate the input data are not entirely helpful in predicting the output. Each feature has a level of association with the predicted output. The goal of GA-KELM is to identify spatiotemporal features and the degree of their association with the predicted output learned through its weight matrix [45].

IV. EXPERIMENTAL RESULTS

In this section, experiments using real-world air quality datasets are conducted to analyze and evaluate the performance of the proposed model. The predictive performance of the model is verified by comparing with other baseline deep learning models.

A. DATASETS DESCRIPTION AND DATA PREPROCESSING

1) DATA COLLECTION AND DATA BASICS

For the study in this paper, we will use 3 real air quality datasets. Our dataset is available on the following link: https://github.com/chunhaoliu/AQI_Data.git. It includes air quality data from July 2019 to July 2021, covering all seasons and different weather conditions. The dataset contains multidimensional air pollutant monitoring data, such as: $SO_2\ (\mu g/m^3)$, $NO_2\ (\mu g/m^3)$, $PM_{10}\ (\mu g/m^3)$, $PM_{2.5}\ (mg/m^3)$, $O_3\ (\mu g/m^3)$ and $CO\ (mg/m^3)$, hourly average concentration data and day-by-day detection data. In the preprocessing stage, the data are initially examined for missing value outliers. To ensure data completeness and accuracy, linear interpolation and 3σ are applied for processing. The original data are then sorted by time series and divide it into a training set and a test set. This dataset will be used to train and test our Improved Extreme Learning Machine (ELM) model. With this approach, we can offer crucial information regarding future air quality conditions to the government and the public. This data-driven insight can assist in the development and implementation of appropriate environmental policies and measures.

2) DATA PREPROCESSING

When dealing with missing data, various methods such as nearest neighbor filling, plural filling, median filling, and mean filling have been widely discussed in the literature. After conducting a thorough literature review and considering the characteristics of the data at hand, the interpolation method is selected as the preferred approach for filling the missing data. This method identifies the closest sample point as the fill value by searching sample points around the missing data. The interpolation method offers advantages such as reduced calculation, high processing efficiency, high flexibility and ease of implementation. Moreover, this approach has minimal impact on the correlation between the data and ensures the integrity of the experimental data. The interpolation effect is shown in FIGURE 5.

However, as observed in the aforementioned figure, the presence of many outliers in the data is still evident. Therefore, outlier processing becomes necessary. To eliminate the outliers in the observed data, statistical analysis methods are usually used. To extract multiple observations and remove outliers, statisitcal criteria such as the Lajda criterion and the Grubbs criterion are commonly employed.

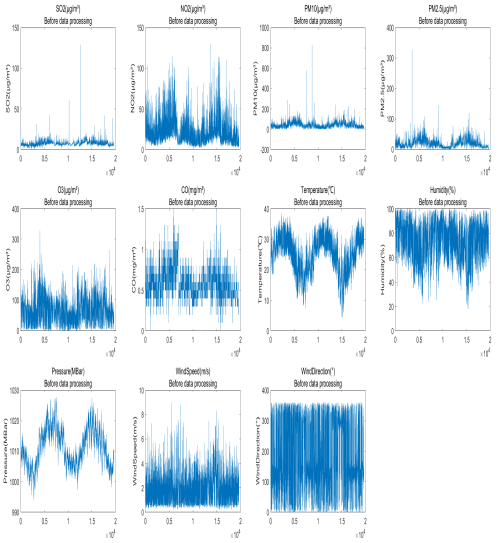

FIGURE 5. Before data processing.

The Lajda criterion may not be highly accurate in removing and identifying outliers when the number of measurements is small (n<10). This method follows a three-step process for determining outliers. First, it assumes that the dataset contains only chance errors. Second, it computes the standard deviation of this set of data, and determines the interval range by certain probability. As data within the range of (μ-3σ, μ+3σ) account for 99.74% of the total data, the probability of exceeding this range is less than 0.3%. Therefore, the distribution characteristics of the vast majority of the data align with the 3σ criterion, conforming to a normal distribution.The standard deviation was calculated as follows:

$$S_g = \sqrt{\frac{\sum_l^n \left(P_i - \bar{P}\right)^2}{n-1}} \tag{16}$$

where P_i is the i th observation, X_g is the standard deviation of the observation, $\bar{P}$ is the average of n observations, and n is the number of observations. For a suspicious data X_i, if $V_i = \left|X_i - \bar{X}\right| > 3S_g$, then X contains coarse deviation and is rejected. The data are then processed using MATLAB after using outliers, as shown in FIGURE 6, and they were obviously smoothed.

B. EVALUATION METRICS

The GA-KELM framework is compared with three baseline deep learning models: CMAQ, SVR, and DBN-BP. CMAQ (Community Multiscale Air Quality) is a computer model used to simulate and predict atmospheric pollutant concentrations and air quality. SVR (Support Vector Regression) is a machine learning algorithm used to model and predict regression problems. It is an extension based on the support vector machine (SVM) algorithm designed to solve

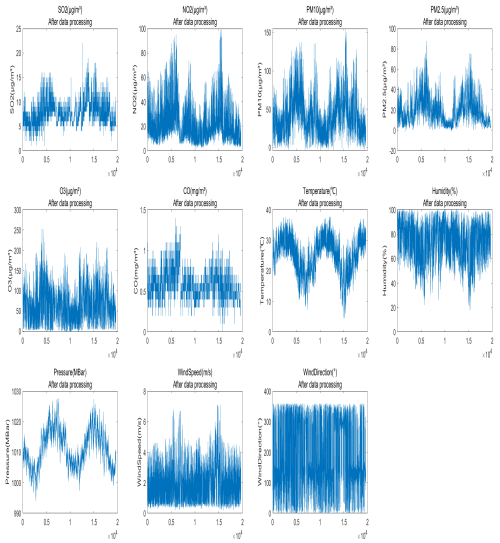

FIGURE 6. After data processing.

regression problems. DBN-BP (Deep Belief Network with Backpropagation) is a neural network model that combines deep confidence network (DBN) and backpropagation (BP) algorithms. In the experiments, 70% of the data were used for training and 30% for testing. The predictive performance of the proposed model was assessed using several evaluation metrics, such as RMSE, mean square error (MSE), and correlation coefficient (R^2).

RMSE quantifies the difference between predicted and actual data. A low value of RMSE underlines a good forecast.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}\left(X_i - \hat{X}_i\right)^2}{N}} \tag{17}$$

MSE is calculated as the squared difference between the true value and the predicted value, which is then averaged. It is defined over the range [0,+∞), where it is 0 when the predicted value is exactly the same as the true value, and the larger the error, the larger the value.

$$\mathrm{MSE} = \frac{1}{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{n}}\left(\hat{\mathrm{y}}_{\mathrm{i}} - \mathrm{y}_{\mathrm{i}}\right)^2 \tag{18}$$

R^2 is a correlation coefficient that defines how well a predicted value matches the original one. The higher the value of R^2, the better the predictive model. R^2 is given by:

$$R^2 = 1 - \frac{\sum\left(X_i - \hat{X}_i\right)^2}{\sum\left(X_i - \bar{X}\right)^2} \tag{19}$$

Air quality is measured using the AQI obtained by calculating the concentration of each pollutant being monitored. The classification of pollutants in the AQI can help in understanding the status of air pollutants. From the measured concentration values, the AQI classification range for each pollutant is strictly determined according to the Ambient AQI, as shown in Table 2.The prediction performance of GA-KELM is assessed by calculating the accuracy of the predicted values in the actual categories.

C. CORRELATION WITHIN THE METEOROLOGICAL PARAMETERS

This section examines the correlation between variables to make more accurate predictions. When pollutant sources are identified, pollutant concentrations are basically determined by meteorological conditions without analyzing the changes brought by the pollutant emissions to the AQI. Therefore, a change in pollutant concentration inevitably causes a change in AQI, and meteorological factors in turn determine the pollutant concentration. Meteorological factors have a significant role in AQI change. The main meteorological factors that affect the AQI are climatic conditions, weather conditions, temperature, wind, humidity, and precipitation. Temperature is one of the most basic meteorological factors among them. Higher temperatures increase the water vapor content in the air. Increased wind speed accelerates air flow, making it easier for pollutants to be transported to higher altitudes, which in turn causes increased pollution at ground level. Relative humidity is related to the moisture in the air. Turbulence in the atmosphere has a diffusion effect on pollutants, causing a decrease in the AQI in the region. Furthermore, temperature and humidity significantly affect the pollutant chemical reaction process. The wind direction affects the diffusion of pollutants, which in turn causes changes in pollutant concentrations and the AQI.

Therefore, for the impact of meteorological factors on AQI, the correlation between the five variables of temperature, humidity, air pressure, wind speed, wind direction and the monitored concentrations of SO_2, NO_2, CO, PM_{10}, $PM_{2.5}$, and O_3 can be considered. Typically, to study the correlation between two groups of variables $(x_1, x_2, x_3, \cdots, x_p)$, $(y_1, y_2, y_3, \cdots y_q)$, the $p * q$ correlation coefficients between the two sets of variables can be calculated, using an idea similar to that of principal components. A certain linear combination of each of the two sets of variables is identified and the correlation between the linear combinations is discussed. In order to have maximum correlation, it is necessary to find the first pair of linear combinations in each set of variables:

$$\begin{cases} u_1 = a_{11}x_1 + a_{21}x_2 + \ldots a_{p1}x_p \\ v_1 = b_{11}y_1 + b_{21}y_2 + \ldots b_{q1}y_q \end{cases} \tag{20}$$

Similarly, the second pair of linear combinations with the next largest correlation is found, so that they are respectively uncorrelated with the first linear combination pair within this group:

$$\begin{cases} u_2 = a_{12}x_1 + a_{22}x_2 + \ldots a_{p2}x_p \\ v_2 = b_{12}y_1 + b_{22}y_2 + \ldots b_{q2}y_q \end{cases} \tag{21}$$

where u_2 and v_2 are correlated, and u_1 is uncorrelated with u_2. In this way, r steps are carried out until the correlation between the two groups of variables is extracted,

TABLE 1. Correlation of various pollutants with five meteorological conditions.

AQI Category Range	$SO_2(\mu g/m^3)$	$NO_2(\mu g/m^3)$	$PM_{10}(\mu g/m^3)$	$PM_{2.5}(\mu g/m^3)$	$O_3(\mu g/m^3)$	$CO(mg/m^3)$
Temperature($^\circ C$)	30.13	25.82	28.43	31.72	18.55	17.73
Humidity(%)	15.72	-3.88	0.89	-0.95	29.32	6.83
Pressure(MBar)	-7.17	-2.69	-1.99	1.43	-27.25	20.19
Wind speed(m/s)	-2.65	-11.08	-5.16	-0.68	-4.39	-2.81
Wind direction(ř)	10.58	-4.14	1.15	-0.72	4.36	-13.48

and the r group of variables can be obtained, with $r \leq \min(p, q)$. To study the correlation between two sets of random variables, the full correlation coefficient is used. The correlation model selects temperature, humidity, barometric pressure, wind speed, and wind direction as meteorological factor indicators, assuming that the correspondence between NO_2 and SO_2 (fuel combustion) is not taken into account. $SO_2, NO_2, CO, PM_{10}, PM_{2.5}$, and O_3 are selected as the source indicators in the model, and all factors are selected based on daily average values. The influence of each weather factor on each pollutant is divided into positive and negative correlations through separate statistical correlations.

Table 1 shows that the atmospheric water-holding capacity is enhanced with temperature rise in typical times, the atmospheric movement slows, the accumulation rate of pollutants exceeds the diffusion rate, and pollutants are not easily diffused, leading to an increase in pollutant concentration and in AQI. The increase in humidity reflects the enhanced water holding capacity of the atmosphere, which also makes the accumulation rate of pollution higher than the diffusion rate, thus causing the pollutant concentration to increase and the AQI to rise. However, increased humidity can also be caused by rainfall that can decrease pollutant concentrations, and consequently, AQI is decreased. Pressure affects air quality mainly as a result of seasonal changes rather than day-to-day variations. For instance, low-pressure weather dominates in winter and the atmosphere moves slowly, which is unfavorable to pollutant dispersion. Wind speed is consistent with wind direction. The analysis of the relationship between meteorological factors and pollutants reveals that the total amount of pollutant emissions varies greatly under different meteorological conditions, and pollutant emissions decrease with an increasing wind speed under the same meteorological conditions. Conceptually, an increase in wind speed causes a decrease in pollution concentration and a decrease in AQI at the designated monitoring stations. However, the wind speed increases in some direction whereas the AQI decreases at the monitoring stations.

D. POLLUTANT CONCENTRATION FORECASTING RESULTS ANALYSIS

In the experiment, a GA is used to optimize the parameters of KELM. The fitness calculated by the GA is a weighted combination of AQI and primary pollutant error. Data from five previous hours (actual collected values including pollutants and weather) are used as input. For instance, assuming that the moment to be predicted is t, then data at $t-1$, $t-2$, $t-3$, $t-4$, and $t-5$ are taken as input. The output is the pollutant concentration to be predicted at this moment (i.e., at time t), as shown in FIGURE 7.

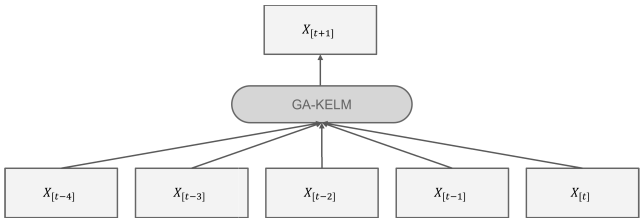

FIGURE 7. End-to-end architecture of GA-KELM.

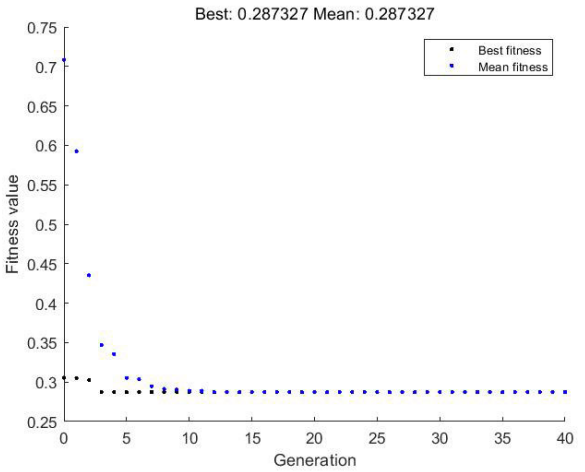

FIGURE 8. Results of the change in fitness for GA optimization.

In the GA, the number of iterations is set to 100, the population size is 10, the number of hidden layer neurons is 30, the number of output layer neurons is 1 representing the predicted AQI, the number of training iterations is 1000, the learning rate is 0.05, and the training error target is set to 0.001. Additionally, the crossover probability is varied in the range [0.6 1.0] in increments of 0.05, and the variance probability is varied in the range [0 1] with an increment of 0.01. RMSE is chosen as the fitness parameter for the optimization problem. For each parameter set, 20 independent replications are performed. For each case, the RMSE values converge rapidly to a stable point.

The parameters to be optimized, in this case the penalty factor and the kernel parameters, are first set at random. The data are divided into two parts: training and prediction of the remaining days. The training samples and the set parameters are used to train the model, and the test samples are used to produce prediction results for AQI and major pollutants. A weighted combination of the maximum relative

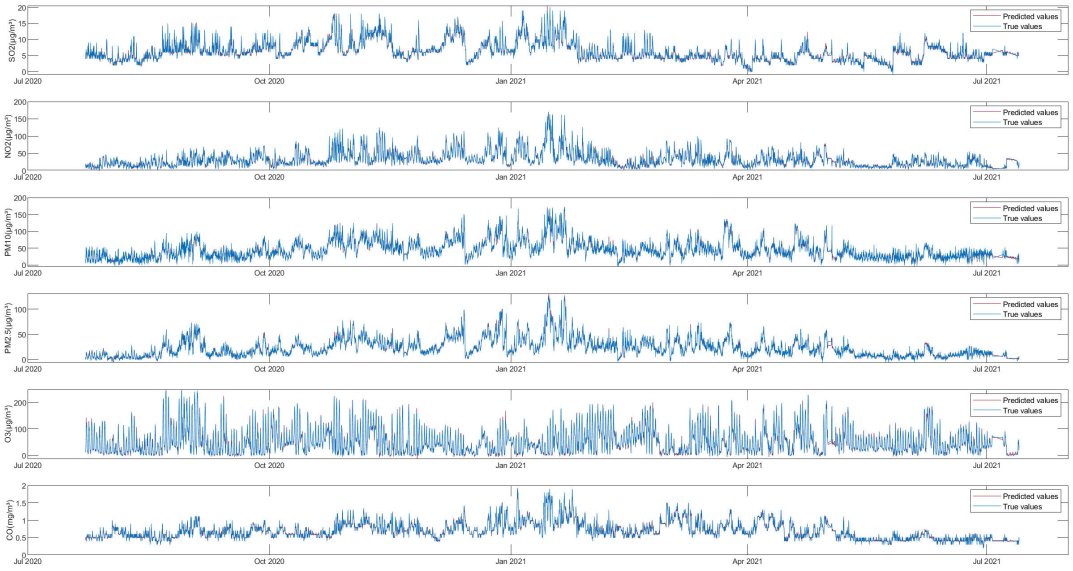

FIGURE 9. Annual air quality forecast comparison (Dataset A).

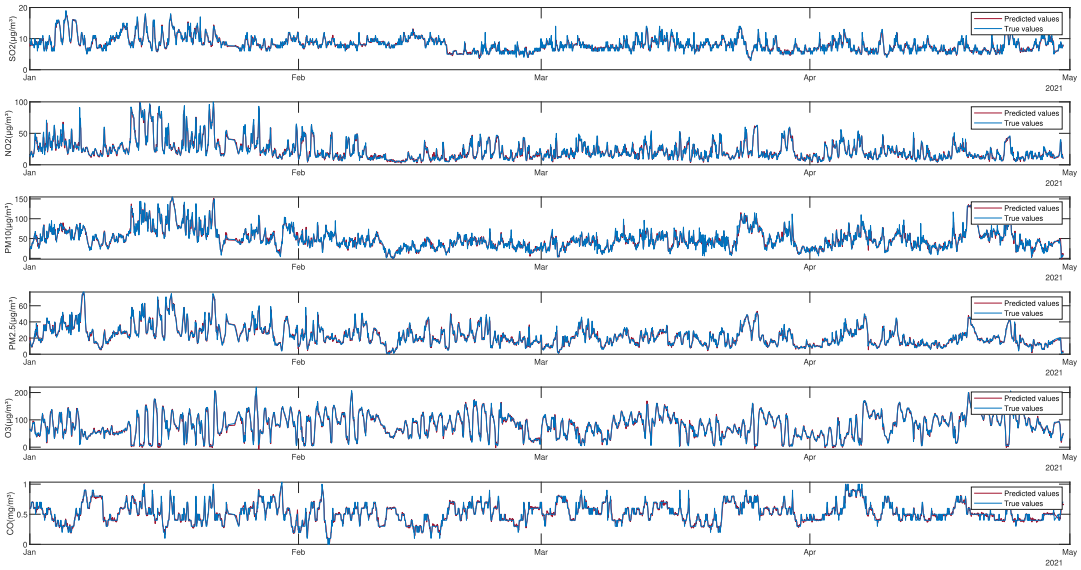

FIGURE 10. Seasonal air quality forecast trends (Dataset B).

TABLE 2. Category range of air pollutants (technical regulation on ambient air quality index, HJ633-2012).

AQI Category Range	$IQAI$	CO	SO_2	NO_2	O_3	PM_{10}	$PM2.5$
Good	0-50	0-2	0-50	0-40	0-100	0-50	0-35
Satisfactory	51-100	3-4	51-150	41-80	101-160	51-150	36-75
Moderately Polluted	101-150	5-14	151-475	81-180	161-215	151-250	76-115
Poor	151-200	15-24	476-800	181-280	216-265	251-350	116-150
Very Poor	201-300	25-36	801-1600	281-565	266-800	351-420	151-25
Severe	300+	36+	1600+	565+	800+	420+	250+

error of AQI and the average relative error of major pollutants is used as our fitness function. Then the GA is used to optimize the fitness function, and the penalty factor and kernel parameter are used as the optimization parameters. According to previous theory, the standard GA will complete convergence at an early stage. The maximum evolutionary generation is set to 40, and the kernel function selected in KELM is the Gaussian RBF. The fitness curve of GA optimization is shown in Fig.8. It can be observed that the fitness keeps decreasing and finally reaches convergence. At this point, the optimal penalty factor and kernel function parameters can be obtained.

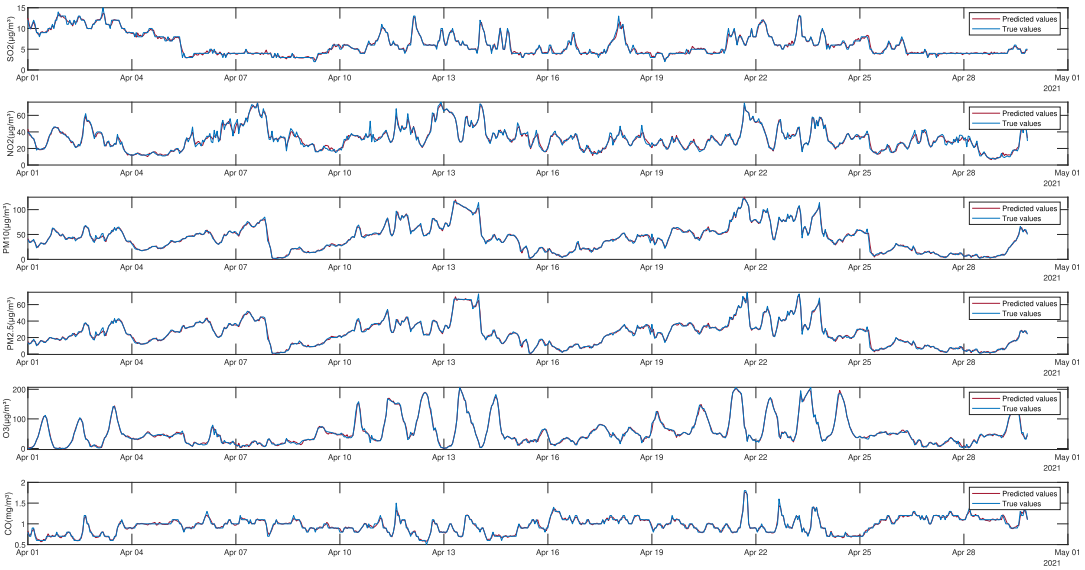

FIGURE 11. Monthly air quality forecast accuracy (Dataset C).

TABLE 3. Comparison of RMSE of GA-KELM with other baseline approaches.

Model	SO_2	NO_2	PM_{10}	$PM_{2.5}$	O_3	CO
CMAQ	41.50	24.19	65.85	36.39	60.58	0.56
SVR	30.75	20.46	44.79	28.18	50.88	50.78
DBN-BP	11.18	14.63	22.25	11.18	17.73	0.25
GA-KELM	9.42	11.23	20.34	6.78	10.34	0.21

TABLE 4. Comparison of MSE of GA-KELM with other baseline approaches.

Model	SO_2	NO_2	PM_{10}	$PM_{2.5}$	O_3	CO
CMAQ	25.63	19.07	8.22	25.63	11.63	6.95
SVR	18.36	18.87	19.76	21.38	34.67	34.67
DBN-BP	8.22	11.63	14.63	8.22	12.86	0.19
GA-KELM	6.73	9.24	10.54	6.78	11.93	0.56

TABLE 5. Comparison of R2 of GA-KELM with other baseline approaches.

Model	SO_2	NO_2	PM_{10}	$PM_{2.5}$	O_3	CO
CMAQ	0.69	0.58	0.87	0.69	0.86	0.89
SVR	0.72	0.63	0.67	0.77	0.78	0.78
DBN-BP	0.87	0.86	0.82	0.87	0.9	0.93
GA-KELM	0.89	0.91	0.85	0.91	0.92	0.95

The obtained optimal parameters are then used to train the limit learning machine. The actual values are compared with the predictions of GA-KELM from July 2020 to July 2021; the red line represents the predicted values, and the blue color represents the actual values. The results of three different datasets A, B, and C are shown in FIGURE 9, FIGURE 10, and FIGURE 11, respectively. As can be seen from the figures, the training effect is relatively good.

In China, the National Air Quality Index sets breakpoints for all air pollutants under 6 categories, namely good, satisfactory, moderately polluted, poor, extremely poor, and severe, as shown in Table 2. During the validation process, the predicted values are grouped under each AQI category based on the predicted value. The ratio of correct predictions to total predictions is estimated to measure the accuracy of the model in predicting the correct AQI category.

The performance of the model in predicting air pollutant concentrations is measured based on the RMSE, MSE and R_2 values of all predicted pollutants. The results are compared with baseline methods for predicting air pollutant concentrations, such as CMAQ, SVR, and DBN-BP, with meteorological parameters as auxiliary inputs. However, the factors considered in model evaluation are different.

The proposed and baseline models were initially trained using the training dataset, and from Table3-5, we infer that GA-KELM air has smaller RMSE and MSE values and higher R_2 values compared to baseline methods for predicting concentrations of various air pollutants.The reduced RMSE and MSE values elucidate the reliability of the GA-KELM air predictions, while the reduced RMSE and improved R_2 values infer the specificity of its mean predictions. The superior performance of GA-KELM illustrates its efficiency in accurately capturing spatiotemporal relationships and their impact on predicted values.

V. 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, SO_2, and PM_{10}.

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] 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 PM_{10} 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 ($PM_{2.5}$) 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 Fe_2O_3/Mn_3O_4 magnetic nanocomposite for enhanced adsorption of methyl orange dye: Neural network modeling and response surface methodology optimization," *Appl. Organometallic Chem.*, vol. 32, no. 3, p. e4186, Mar. 2018.

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

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

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

[14] S. Du, T. Li, Y. Yang, and S. Horng, "Deep air quality forecasting using hybrid deep learning framework," *IEEE Trans. Knowl. Data Eng.*, vol. 33, no. 6, pp. 2412–2424, Jun. 2021.

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

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

[17] S. Mahajan, H.-M. Liu, T.-C. Tsai, and L.-J. Chen, "Improving the accuracy and efficiency of $PM_{2.5}$ forecast service using cluster-based hybrid neural network model," *IEEE Access*, vol. 6, pp. 19193–19204, 2018.

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

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

[20] A. Agarwal and M. Sahu, "Forecasting $PM_{2.5}$ concentrations using statistical modeling for Bengaluru and Delhi regions," *Environ. Monit. Assessment*, vol. 195, p. 502, Mar. 2023.

[21] D. J. Lary, T. Lary, and B. Sattler, "Using machine learning to estimate global $PM_{2.5}$ for environmental health studies," *Environ. Health Insights*, vol. 9, no. 1, pp. 41–52, 2015.

[22] Y. Zheng, X. Yi, M. Li, R. Li, Z. Shan, E. Chang, and T. Li, "Forecasting fine-grained air quality based on big data," in *Proc. 21th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, New York, NY, USA, Aug. 2015, pp. 2267–2276.

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

[24] T. S. Rajput and N. Sharma, "Multivariate regression analysis of air quality index for Hyderabad city: Forecasting model with hourly frequency," *Int. J. Appl. Res.*, vol. 3, no. 8, pp. 443–447, 2017.

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

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

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

[28] P. C. Campbell, Y. Tang, P. Lee, B. Baker, D. Tong, R. Saylor, A. Stein, J. Huang, H.-C. Huang, E. Strobach, J. McQueen, L. Pan, I. Stajner, J. Sims, J. Tirado-Delgado, Y. Jung, F. Yang, T. L. Spero, and R. C. Gilliam, "Development and evaluation of an advanced national air quality forecasting capability using the NOAA Global Forecast System version 16," *Geosci. Model Develop.*, vol. 15, no. 8, pp. 3281–3313, Apr. 2022.

[29] X.-B. Jin, Z.-Y. Wang, W.-T. Gong, J.-L. Kong, Y.-T. Bai, T.-L. Su, H.-J. Ma, and P. Chakrabarti, "Variational Bayesian network with information interpretability filtering for air quality forecasting," *Mathematics*, vol. 11, no. 4, p. 837, Feb. 2023.
[30] J. Li, W. Shi, and D. Yang, "Color difference classification of dyed fabrics via a kernel extreme learning machine based on an improved grasshopper optimization algorithm," *Color Res. Appl.*, vol. 46, no. 2, pp. 388–401, Apr. 2021.
[31] Z. Zhou, Y. Hu, J. Ji, Y. Wang, Z. Zhu, D. Yang, and J. Chen, "Robot manipulator visual servoing via Kalman filter-optimized extreme learning machine and fuzzy logic," *KSII Trans. Internet Inf. Syst.*, vol. 16, no. 8, pp. 2529–2551, 2022.
[32] Z. Zhou, W. Deng, Z. Zhu, Y. Wang, J. Du, and X. Liu, "Fabric defect detection based on feature fusion of a convolutional neural network and optimized extreme learning machine," *Textile Res. J.*, vol. 92, nos. 7–8, pp. 1161–1182, Apr. 2022.
[33] Y. Xue, Q. Wu, and H. Sun, "Air quality prediction model based on genetic algorithm and weighted extreme learning machine," *Proc. SPIE*, vol. 12294, Oct. 2022, Art. no. 1229454.
[34] X. Liang, G. Wang, M. R. Min, Y. Qi, and Z. Han, "A deep spatio-temporal fuzzy neural network for passenger demand prediction," in *Proc. SIAM Int. Conf. Data Mining (SDM)*, 2019, pp. 100–108.
[35] Y. Liu, C. Lyu, A. Khadka, W. Zhang, and Z. Liu, "Spatio-temporal ensemble method for car-hailing demand prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 12, pp. 5328–5333, Dec. 2020.
[36] J. Ma, Y. Ding, J. C. P. Cheng, F. Jiang, and Z. Wan, "A temporal-spatial interpolation and extrapolation method based on geographic long short-term memory neural network for $PM_{2.5}$," *J. Cleaner Prod.*, vol. 237, Nov. 2019, Art. no. 117729.
[37] J. Ma, Y. Ding, V. J. L. Gan, C. Lin, and Z. Wan, "Spatiotemporal prediction of $PM_{2.5}$ concentrations at different time granularities using IDW-BLSTM," *IEEE Access*, vol. 7, pp. 107897–107907, 2019.
[38] S. Maji, S. Ghosh, and S. Ahmed, "Association of air quality with respiratory and cardiovascular morbidity rate in Delhi, India," *Int. J. Environ. Health Res.*, vol. 28, no. 5, pp. 471–490, Sep. 2018.
[39] M. Wang, H. Chen, B. Yang, X. Zhao, L. Hu, Z. Cai, H. Huang, and C. Tong, "Toward an optimal kernel extreme learning machine using a chaotic moth-flame optimization strategy with applications in medical diagnoses," *Neurocomputing*, vol. 267, pp. 69–84, Dec. 2017.
[40] M. A. E. Aziz, A. A. Ewees, and A. E. Hassanien, "Whale optimization algorithm and moth-flame optimization for multilevel thresholding image segmentation," *Expert Syst. Appl.*, vol. 83, pp. 242–256, Oct. 2017.
[41] Y. Yang, Y. Wang, and X. Yuan, "Parallel chaos search based incremental extreme learning machine," *Neural Process. Lett.*, vol. 37, no. 3, pp. 277–301, Jun. 2013.
[42] F. Jiang, J. He, and Z. Zeng, "Pigeon-inspired optimization and extreme learning machine via wavelet packet analysis for predicting bulk commodity futures prices," *Sci. China Inf. Sci.*, vol. 62, no. 7, pp. 1–19, Jul. 2019.
[43] Q. Li, H. Chen, H. Huang, X. Zhao, Z. Cai, C. Tong, W. Liu, and X. Tian, "An enhanced grey wolf optimization based feature selection wrapped kernel extreme learning machine for medical diagnosis," *Comput. Math. Methods Med.*, vol. 2017, Jan. 2017, Art. no. 9512741.
[44] G.-B. Huang, D. H. Wang, and Y. Lan, "Extreme learning machines: A survey," *Int. J. Mach. Learn. Cybern.*, vol. 2, no. 2, pp. 107–122, Jun. 2011.
[45] G. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 42, no. 2, pp. 513–529, Apr. 2012.

CHUNHAO LIU (Graduate Student Member, IEEE) was born in Linyi, Shandong, China, in 1997. He received the B.S. degree in automation from Linyi University, China, in 2019, where he is currently pursuing the Graduate degree in electronics and information technology with the School of Automation and Electrical Engineering. His current research interests include machine learning and deep learning, data mining, and control systems.

GUANGYUAN PAN (Member, IEEE) was born in Linyi, Shandong, China, in 1987. He received the B.S. degree in automation from the Qingdao University of Science and Technology, China, in 2009, and the M.S. and Ph.D. degrees in control science and engineering from the Beijing University of Technology (Dr. Junfei Qiao's team), Beijing, China, in 2012 and 2016, respectively. From 2013 to 2014, he was a joint-training Ph.D. student with the University of Victoria (Dr. Nikitas Dimopoulos' team), BC, Canada. He was a Postdoctoral Researcher with the Department of Civil and Environmental Engineering, University of Waterloo, ON, Canada, from 2016 to 2019. He is currently an Associate Professor with the Department of Automation and Electrical Engineering, Linyi University, China. His current research interests include machine learning and deep learning, data mining and signal processing, transportation safety, and intelligent systems.

DONGMING SONG was born in Yantai, Shandong, China, in 1976. He received the bachelor's degree in mechanical design and manufacturing from the China University of Petroleum (East China), in 1999, and the master's degree in mechanical and electronic engineering and the Ph.D. degree in computer application technology from the Nanjing University of Science and Technology, in 2004 and 2009, respectively. He is currently an Associate Professor with the Department of Automation and Electrical Engineering, Linyi University, China. His current research interests include energy internet, energy big data, electrical, and control systems.

HAO WEI was born in Linyi, Shandong, China, in 1989. He received the bachelor's degree in computer science from Imperial College London, U.K., in 2011, and the master's degree in software engineering and the Ph.D. degree in computer science from the University of York, U.K., in 2013 and 2018, respectively. He is currently a Lecturer with the Department of Automation and Electrical Engineering, Linyi University, China. His current research interests include swarm robotics, genetic algorithm, and system development related to artificial intelligence and big data.