Predicting air quality index using attention hybrid deep learning and quantum-inspired particle swarm optimization

Anh Tuan Nguyen[1], Duy Hoang Pham[1], Bee Lan Oo[2], Yonghan Ahn[1*] and Benson T. H. Lim[2]

*Correspondence:
yhahn@hanyang.ac.kr

[1] Smart City Engineering,
Hanyang University ERICA
Campus, 55 Hanyangdeahak-ro,
Sangnok-gu, Ansan,
Gyeonggi-do 15588, South Korea
[2] School of Built Environment,
University of New South Wales,
Kensington, Sydney, NSW 2052,
Australia

Abstract

Air pollution poses a significant threat to the health of the environment and human well-being. The air quality index (AQI) is an important measure of air pollution that describes the degree of air pollution and its impact on health. Therefore, accurate and reliable prediction of the AQI is critical but challenging due to the non-linearity and stochastic nature of air particles. This research aims to propose an AQI prediction hybrid deep learning model based on the Attention Convolutional Neural Networks (ACNN), Autoregressive Integrated Moving Average (ARIMA), Quantum Particle Swarm Optimization (QPSO)-enhanced-Long Short-Term Memory (LSTM) and XGBoost modelling techniques. Daily air quality data were collected from the official Seoul Air registry for the period 2021 to 2022. The data were first preprocessed through the ARIMA model to capture and fit the linear part of the data and followed by a hybrid deep learning architecture developed in the pretraining–finetuning framework for the non-linear part of the data. This hybrid model first used convolution to extract the deep features of the original air quality data, and then used the QPSO to optimize the hyperparameter for LSTM network for mining the long-terms time series features, and the XGBoost model was adopted to fine-tune the final AQI prediction model. The robustness and reliability of the resulting model were assessed and compared with other widely used models and across meteorological stations. Our proposed model achieves up to 31.13% reduction in MSE, 19.03% reduction in MAE and 2% improvement in R-squared compared to the best appropriate conventional model, indicating a much stronger magnitude of relationships between predicted and actual values. The overall results show that the attentive hybrid deep Quantum inspired Particle Swarm Optimization model is more feasible and efficient in predicting air quality index at both city-wide and station-specific levels.

Keywords: Air Quality Index, Long short term memory, Time series forecasting, Quantum particle swarm optimization, Attention, Hybrid deep learning

Introduction

Air pollution is a severe global issue due to its detrimental impact on human health and the environment. It is particularly prominent in certain regions, including South Korea, where it poses a significant threat, ranging from respiratory ailments to severe illnesses

Nguyen *et al. Journal of Big Data*   (2024) 11:71

Page 2 of 38

such as cancer, heart diseases, and cardiovascular complications. The introduction of hazardous or excessive levels of substances like gases, particles, and biological molecules into the Earth's atmosphere results in air pollution. Pollutants and fine particulate matter (PM) that contribute to air pollution include nitrogen dioxide (NO_2), carbon monoxide (CO), carbon dioxide (CO_2), ozone (O_3) and sulfur dioxide (SO_2) [1]. The rapid industrialization, urbanization, and transportation of South Korea have caused severe air pollution, resulting in the release of pollutants in the air and greenhouse gases [2]. Particularly, Seoul, one of the densely populated metropolises in the world, faces numerous air pollution challenges due to its high-intensity industries, automobile emissions, and meteorological conditions. The city's rapid growth and urbanization have led to a concentration of factories and vehicles, which release significant amounts of pollutants into the atmosphere [3]. Moreover, Seoul's geography, nestled between mountains and the ocean, can trap pollutants and worsen air quality.

According to Zou et al. [4], air pollution poses a significant threat to public health, prompting widespread concern about future air quality trends. It is associated with a range of adverse health effects, including asthma, weakened lung function, increased cardiopulmonary illnesses, and elevated mortality rates. To address these challenges, Seoul government has implemented various strategies to curb air pollution. These include promoting public transportation, encouraging the use of cleaner fuels, investing in renewable energy sources, establishing air quality monitoring networks and implementing public awareness campaigns to educate citizens about the risks of air pollution. While the progress has been encouraging, air pollution remains a significant concern in Seoul. Particularly, the report by World Health Organization has shown that, Seoul has an extremely poor PM concentration level ($46\,\mu g/m^3$) compared with major cities across the world [5]. It follows that the city government is facing the challenges of actively maintaining and improving its economic growth, and continually reducing carbon emissions, improving air quality, and ensuring the health and well-being of the Seoul's residents. This further add weights to Koo et al.'s [6] recommendations of a real-time air quality monitoring and prediction system towards supporting urban planners, policy makers, and air quality agencies in implementing sustainable development strategies.

Table 1 shows the Korean air quality index (AQI) levels and their corresponding pollutant concentrations and health impacts [7]. The AQI is a numerical measure that quantifies the air quality in a given location, including the concentrations of multiple pollutants [4]. It serves as a tool for assessing the potential health risks associated with air pollution and evaluating the effectiveness of air quality management strategies. The AQI is calculated based on the latest ambient air quality standards (GB3095-2012), which encompass six key pollutants: ozone (O_3), carbon monoxide (CO), nitrogen dioxide (NO_2), sulfur dioxide (SO_2), respirable particulate matter (PM_{10}) and fine particulate matter ($PM_{2.5}$).

AQI prediction methods for air pollutants may be classified into four categories: (i) statistics model, (ii) ML-based method, (iii) DL-based methods and (iv) hybrid methods.

Statistical models rely on assumptions about the underlying data distribution to establish causal relationships and heavily emphasize the estimation of unknown parameters. The use of statistical methods to predict air quality primarily involves the autoregressive

Nguyen *et al. Journal of Big Data*   (2024) 11:71

Page 3 of 38

Table 1 Korean AQI levels

AQI level	PM$_{2.5}$ (µg/m^3)	PM$_{10}$ (µg/m^3)	NO$_2$ (ppm)	SO$_2$ (ppm)	O$_3$ (ppm)	CO (ppm)	Health effects
Good (0–50)	0–15	0–30	0–0.03	0–0.02	0–0.03	0–2	A level that will not impact patients suffering from diseases related to air pollution
Moderate (51–100)	16–35	31–80	0.031–0.06	0.021–0.05	0.031–0.09	2.1–9	A level which may have a meager impact on patients in case of chronic exposure
Unhealthy (101–250)	36–75	81–150	0.061–0.2	0.051–0.15	0.091–0.15	9.1–15	A level that may have harmful impacts on patients and members of sensitive groups (children, aged or weak people), and also cause the general public unpleasant feelings
Very unhealthy (251–500)	76–500	151–600	0.201–2	0.151–1	0.151–0.6	15.1–50	A level which may have a serious impact on patients and members of sensitive groups in case of acute exposure, and that even the general public can be weakly affected

(AR) model, the autoregressive integrated moving average (ARIMA) model, the gray model, and the multiple linear regression (MLR) model. Carbajal-Hernández et al. [8] proposed an approach that involves developing an algorithm to assess the pollution level of air quality parameters and creating a new air quality index based on a fuzzy reasoning system. Zhang et al. [9] compared two distinct approaches to model development, including generalized additive models (GAMs) and conventional linear regression techniques. Zhao et al. [10] proposed an ARIMA model for PM2.5 annual data, utilizing the augmented Dickey-Fuller test to demonstrate the need for first-order differencing. Additionally, a seasonally nonlinear gray model was developed to capture the seasonal variations in the time series of seasonally fluctuating pollution indicators, ensuring accurate predictions that effectively capture both seasonal and nonlinear patterns [11].

Machine learning models overcome convergence obstacles and enhance their predictive power by harnessing the insights gleaned from vast amounts of data, enabling

Nguyen *et al. Journal of Big Data*   (2024) 11:71

Page 4 of 38

them to make accurate predictions about future events. Mehmood et al. [12] discussed how conventional methods were transformed into machine learning approaches. Through machine learning-driven analysis of emerging trends, this approach identifies promising research avenues with potential for significant impact. Usha et al. [13] used two machine learning algorithms, which are Neural Networks and Support Vector Machines (SVMs), showed improvement of the prediction accuracy and suggest that the model can be used in other smart cities as well. Elsheikh [14] discussed the applications of machine learning (ML) in friction stir welding (FSW) for predicting joint properties, enabling real-time control, and diagnosing tool failures. Ke et al. [15] proposed a machine learning-based air quality forecasting system to predict daily concentrations of six pollutants using meteorological data, pollutant emissions, and model reanalysis data. The system integrates five machine learning models and automatically selects the best model and hyperparameters. Zhang et al. [16] employed machine learning techniques to anticipate how indoor mode's unpredictability and variability could lead to suboptimal air quality. Gu et al. [17] proposed a new hybrid interpretable predictive machine learning model for the $PM_{2.5}$ prediction, which demonstrated the superiority over other models in prediction accuracy for peak values and model interpretability. Most recently, Rakholia et al. [18] constructed a comprehensive model that incorporated various factors influencing air quality, including meteorological conditions, traffic patterns, levels of air pollution in residential and industrial zones, urban spatial data, time series analysis, and pollutant concentrations.

Building upon the foundation of machine learning (ML) techniques for AQI prediction, it appears necessary here to describe the benefits and application of appropriate types of deep learning (DL) approaches. DL's ability to handle large datasets and achieve superior accuracy has propelled its popularity in AQI prediction. Due to its inherent adaptability and transformability, DL models can be readily adapted to a wide range of domains and applications, surpassing the capabilities of traditional machine learning models. Janarthanan et al. [19] proposed Support Vector Regression (SVR) and LSTM based deep learning model to predict the AQI values accurately and help to plan the metropolitan city for sustainable development. The expected AQI value can control the pollution level by incorporating road traffic signal coordination, encouraging the people to use public transportation, and planting more trees on some locations [19]. Zhang et al. [20] investigated current DL methods for air pollutant concentration prediction from the perspectives of temporal, spatial and spatio-temporal correlations these methods could model. Saez et al. [21] presented a hierarchical Bayesian spatiotemporal model that allowed to make fairly accurate spatial predictions of both long-term and short-term exposure to air pollutants with a relatively low density of monitoring stations and at a much lower computation time. Jurado et al. [22] harnessed the power of convolutional neural networks to create a swift and precise air pollution forecasting system that leverages real-time data on wind speed, traffic flow, and building geometry. Zhou et al. [23] employed a CNN-Gated Recurrent Unit (GRU) model, where the CNN extracted relevant features from the input data and the GRU modeled the temporal dependencies between these features, to predict AQI values. Mao et al. [24] developed a DL framework, a temporal sliding LSTM extended model (TS-LSTME), to predict air quality in the next 24 h using a temporal sliding LSTM

Nguyen *et al. Journal of Big Data*   (2024) 11:71

Page 5 of 38

model that incorporates historical PM2.5 data, meteorological data and temporal data. Elsheikh et al. [25] explored the use of a Long Short-Term Memory (LSTM) neural network to predict the freshwater production of a stepped solar still with a corrugated absorber plate, comparing its performance to a conventional design. Djouider et al. [26] investigated the use of machine learning, specifically LSTM networks and a special relativity search algorithm, to model the effects of friction stir processing on AA2024/ Al_2O_3 nanocomposites, alongside experimental validation.

To further enhance prediction accuracy, researchers have created combinatorial models that enhance prediction rates. By leveraging the strengths of various models, the combined model's prediction accuracy has been significantly elevated. Wu and Lin [27] proposed a novel optimal-hybrid model called SD-SE-LSTM-BA-LSSVM that combines secondary decomposition, AI methods, and an optimization algorithm for practical AQI forecasting. Sarkar et al. [28] combined two DL models like LSTM and GRU models to predict the AQI of the environment, which achieves better result in terms of MAE and R^2 than the other existing approaches. Gilik et al. [29] presented hybrid deep learning model that combines the CNN and LSTM networks to predict air pollutant concentrations in multiple locations across a city, using both univariate and multivariate approaches. Existing forecasting methods like multiple linear models, ARIMA, and SVR are seemed inadequate for capturing the nuances of AQI data [30]. Zhu et al. [1] addressed the limitations of existing AQI forecasting methods, two hybrid models (EMD-SVR-Hybrid and EMD-IMFs-Hybrid) are proposed to improve forecasting accuracy. To further improve forecasting accuracy, Chang et al. [31] proposed a hybrid model that leverages stacking based ensemble learning and Pearson correlation coefficient to integrate various forecasting models. Wang et al. [32] added an attention mechanism to the model to improve the prediction accuracy of the LSTM model. Elsheikh et al. [33] proposed a deep-learning model, specifically a long short-term memory (LSTM) network, to forecast confirmed COVID-19 cases, recoveries, and deaths in Saudi Arabia. Dai et al. [34] established five haze hazard risk assessment models by improving the particle swarm optimization (IPSO) light gradient boosting machine (LightGBM) algorithm and a hybrid model combining XGBoost. Saba and Elsheikh [35] applied nonlinear autoregressive artificial neural networks (NARANN) and statistical methods (ARIMA) to analyze and forecast the COVID-19 outbreak within Egypt, providing insights for policymakers to develop short-term response plans.

The rapid advancements in soft computing technologies have paved the way for the development of numerous meta-heuristic algorithms, which provide simple and easily implementable alternatives to improve the accuracy of predictive models. For example, the Multi-Verse Optimizer (MVO) algorithm, which is driven by cosmological concepts (e.g., white holes, black holes and wormholes), has been developed to effectively balance exploration, exploitation, and local search for optimization tasks [36]. To this, Heydari et al. [37] developed a new intelligent hybrid model based on LSTM and MVO algorithm to analyze and predict air pollution in Combined Cycle Power Plants. Next, the Harris-hawks optimization (HHO) algorithms a nature-inspired group intelligence- based optimization algorithm where the purpose is to minimize or maximize an objective function given a constraint [38]. Du et al. [39] introduced a novel multi-objective optimization variant of the

Nguyen *et al. Journal of Big Data*   (2024) 11:71

Page 6 of 38

HHO algorithm in a hybrid model to enhance the accuracy of the PM_{10} and $PM_{2.5}$ predictive models. Also, inspired by the collaborative foraging strategies of natural organisms, Particle Swarm Optimization (PSO) is a population-based optimization algorithm that employs a swarm of particles to explore the solution space and converge towards optimal solutions [40]. Huang et al. [41] proposed a novel back-propagation (BP) neural network-based method for predicting AQI by employing an improved PSO algorithm with inertia weight variation strategies and learning factors. The Cuckoo Optimization Algorithm (COA)is a metaheuristic optimization algorithm that simulates the actions of a male cuckoo occupying a host's nest and a female cuckoo laying eggs randomly to search for the optimal solution [42]. Sun and Sun [43] presented a novel hybrid model based on principal component analysis (PCA) and least squares support vector machine (LSSVM) optimized by cuckoo search in $PM_{2.5}$ concentration prediction. Inspired by the voting process, Trojovsky and Dehghani [44] proposes a new, leader-selecting, stochastic-based optimization algorithm called the Election-Based Optimization Algorithm (EBOA) to effectively address optimization challenges. Abd Elaziz et al. [45] developed a new model for predicting freshwater production in a membrane desalination system by combining a Long-Short Term Memory (LSTM) network with an Election-Based Optimizer (EBO) for optimization. The Dung Beetle Optimizer (DBO) algorithm has been developed to achieve balances global exploration and local exploitation, resulting in fast convergence and accurate solutions [46]. To address the limitations of CNN-LSTM hyperparameter settings, Duan et al. [47] proposed a hybrid approach, combining an ARIMA model for linear data and a CNN-LSTM model for nonlinear data, optimized using the Dung Beetle Optimizer algorithm for improved accuracy.

The review of literature reveals that the inherent non-stationarity of AQI data poses challenges for individual models to fully capture the intricate patterns of data. Previous studies have mainly compared their proposed models to derivatives of those models, providing an incomplete assessment of alternative approaches and limiting the achievable accuracy. To address these limitations, the aim of this research is to develop an integrative DL model, comprising Attention Convolutional Neural Networks (ACNN), QPSO-LSTM and XGBoost, to predict AQI using Seoul as a case study. AQI datasets, characterized by six of the most prominent pollutants, were extracted from the official Seoul Air Registry for model development and validation. The main contributions of this research are summarized below:

- Quantum particle swarm optimization (QPSO) algorithms were adopted to fine tune the LSTM parameters towards reducing redundancy and saving simulation time. Through the improvement of LSTM, the model could capture irregular patterns that may be otherwise ignored. Attention-based CNN (ACNN) can capture global and local dependence that LSTM may not, enhancing the robustness. In our proposed encoder–decoder framework, we adopted a ACNN–QPSO-LSTM structure.
- To address the complex dynamics of AQI data, the proposed model employed a two-stage approach whereby the first stage involved the extraction and linear fitting of data using the ARIMA model, to yielding the predicted values for the linear component. The nonlinear component is extracted from the residual of

data and is subsequently fed into the hybrid DL model, which then generated the predicted values for the nonlinear component.

- The predicted values from the linear and nonlinear components of data are synthesized to generate the final prediction output. The output is obtained through a XGBoost regressor for precise extraction of features and fine-tuning.
- The proposed hybrid model demonstrates consistent superiority across diverse performance metrics (MSE, MAE, and R^2), suggesting its robustness and generalizability compared to other popular models.

Materials and methods

Statistical method

Autoregressive Integrated Moving Average (ARIMA) is a statistical forecasting method used to predict future values of a time series based on its past values. The ARIMA model is a generalization of the autoregressive moving average (ARMA) model, which assumes that the time series data is stationary, meaning that its statistical properties do not change over time. The ARIMA model, on the other hand, can be used to forecast nonstationary time series data by first differencing the data to make it stationary, and the mathematical model can be represented by Eq. (1). The Augmented Dickey–Fuller (ADF) [48] test was applied to both the original and first-order differenced sequences of each pollutant concentration time series to ensure stationarity and guide appropriate time series modeling techniques. When the p-value ≤ 0.01 and the test statistic value $\geq$ the critical value (1%), the sequence is stationary.

$$y_t = c + \phi_1 * y_{t-1} + \cdots + \phi_p * y_{t-p} + \theta_1 * e_{t-1} + \cdots + \theta_q * e_{t-q}. \tag{1}$$

The p and q are determined by the Akaike Information Criterion (AIC), which is a measure of the relative quality of statistical models for a given set of data. It is a widely used measure in time series forecasting, including the selection of the order of an ARIMA model. The AIC is a relative measure, meaning that it can be used to compare different models of the same data. The model with the lowest AIC could considered as the best-fitting model. In the context of ARIMA models, the AIC is calculated as

$$AIC = -2ln(L) + 2(p + q + k + 1), \tag{2}$$

where y_t is the number of difference levels, c is a constant value, ϕ is the AR parameter (autocorrelation size), p is the number of lags (AR order), θ is the MA parameter value (error autocorrelation), q denotes the number of lags (order of the model MA), e_t is the error, k is the number of model parameters, n is the number of samples and L is the likelihood function.

Deep learning model

Long short term memory (LSTM)

Long-short-term memory (LSTM) is a type of recurrent neural network (RNN) that is used to solve the problem of vanishing gradients [49]. This problem occurs when the gradients of the error function become very small or very large during the backpropagation algorithm, which can prevent the network from learning effectively.

Nguyen *et al. Journal of Big Data*   (2024) 11:71

Page 8 of 38

LSTMs are able to overcome this problem by using a special type of cell that has three gates: the forget gate, the memory gate, and the output gate.

The task of the forget gate is to accept a long-term memory C_{t-1} (the output from the previous unit module) and decide which part of C_{t-1} to retain and forget. The input gate is designed to erase the rejected attribute information by the gate, identify the corresponding fresh attribute information in the unit module, and replace the discarded attribute information. The output gate plays a crucial role in determining the output of the cell state. The cell state undergoes processing via the tanh layer, and the resultant values are multiplied to yield the final information for output.

Deep learning sequence model

Unlike recurrent neural networks (RNNs), basic feedforward neural networks (FFNNs) are unable to effectively model time series data due to their exclusive dependence on current input I_t to generate the corresponding output O_t. RNNs address this limitation by introducing a delay mechanism that preserves the latent state H_{t-1} of the previous time step, allowing the network to incorporate temporal context into its output O_t alongside the current input I_t. This ability to account for historical information enhances the network's capacity to model sequential data. As error signals propagate over time in a recurrent neural network (RNN), they can become increasingly small or even zero, making it difficult for the network to learn long-term dependencies. This is known as the vanishing gradient problem. Long Short-Term Memory (LSTM) networks [50] are a type of RNN that is specifically designed to address the vanishing gradient problem. LSTMs use gated activation functions to selectively remember updated information and forget accumulated information. This allows them to learn long-term dependencies in sequential data.

A sequence-to-sequence (seq2seq) model [51] is a type of neural network architecture that utilizes an encoder–decoder structure to analyze and process sequential data. This architecture enhances the ability of Long Short-Term Memory (LSTM) networks to learn hidden information from noisy data. The seq2seq model consists of two main components:

- Encoder: This component is an LSTM network that processes the input sequence and encodes it into a context vector (usually represented by the hidden state at the last time step, h_N).
- Decoder: This component takes the context vector generated by the encoder as input and decodes it to produce an output sequence. In particular, the output from the previous time step is used as the input from the next time step in the decoder.

This encoder–decoder architecture allows the seq2seq model to learn long-term dependencies in the input sequence and generate an output sequence that is relevant to the input. To improve the quality of the decoded sequence in a seq2seq model, a beam search is employed. Both beam search and the Viterbi algorithm used in Hidden Markov models (HMMs) share a foundation in dynamic programming. In HMM decoding, the process of finding the optimal state estimate based on observations and the previous state is known as "inference." This involves solving for the conditional

probability of the current state given the observations up to that point $p(x_k|y_{1:k})$ and the conditional probability of the current state given all observations $p(x_k|y_{1:N})$. These calculations are equivalent to the forward-backward algorithm in HMMs. By combining forward and backward estimates, the optimal bidirectional estimate of the state can be obtained through the distribution of x_k. This probabilistic perspective forms the basis for bidirectional LSTMs, which combine forward and backward information to achieve better performance [52].

Attention mechanism

The attention mechanism has gained significant attention in various fields, particularly in the context of machine translation and neural network architectures, introduced Transformer, a network architecture based solely on attention mechanisms, which eliminates the need for recurrence and convolutions [53]. Inspired by human attention, the attention mechanism of DL models highlight key data points for enhanced performance. For input $X = (x_1, x_2, \ldots, x_k)$, give query vector q, depict the index of selected information by attention $z = 1, 2, \ldots, N$, then the distribution of attention [54]:

$$\alpha_i = \text{softmax}(s(x_i, \mathbf{q})) = \frac{\exp(\text{score}(x_i, \mathbf{q}))}{\sum_{j=1}^{N} \exp(\text{score}(x_j, \mathbf{q}))}. \tag{3}$$

Here,

$$\text{score}(x_i, \mathbf{q}) = \frac{x_i^T \mathbf{q}}{\sqrt{d}} \tag{4}$$

is attention score through scaled dot product [53], d is the dimension of input. Let $(K, V) = [(k_1, v_1), \ldots, (k_N, v_N)]$ represent the input key-value pairs. The attention function, with a specific query q, is described below:

$$\text{Attention}(K, V), \mathbf{q} = \sum_{i=1}^{N} \alpha_i \mathbf{v}_i = \sum_{i=1}^{N} \frac{\exp(\text{score}(\mathbf{k}_i, \mathbf{q}))}{\sum_j \exp(\text{score}(\mathbf{k}_j, \mathbf{q}))}. \tag{5}$$

Multi-head mechanism is adopted through multi-query $Q = [\mathbf{q}_1, \ldots, \mathbf{q}_M]$ for attention function computation. Multi-head attention (MHA) function is described, see from [54]:

$$\text{Attention}((K, V), Q) = \text{Attention}((K, V), \mathbf{q}_1 \| \ldots \| \text{Attention}((K, V), \mathbf{q}_M)). \tag{6}$$

Here, || denotes Concatenate operation.

The attention mechanism can be employed to learn data-driven weights represented by Q, K, V. These are obtained through linear transformations of X with matrices W_Q, W_K, W_V, respectively, which can be dynamically updated during training.

$$\begin{aligned} Q &= W_Q X \\ K &= W_K X. \\ V &= W_V X \end{aligned} \tag{7}$$

This is called self-attention. Similarly, output

$$\mathbf{h}_i = \text{Attention}((K, V), \mathbf{q}_i). \tag{8}$$

Hence

$$\mathbf{h}_i = \sum_{j=1}^{N} \alpha_{ij} \mathbf{v}_j = \sum_{j=1}^{N} \text{softmax}(s(\mathbf{k}_j, \mathbf{q}_i)) \mathbf{v}_j. \tag{9}$$

Adopting scaled dot product score, the output is

$$H = \text{softmax}\left(\frac{QK^T}{\sqrt{d}}\right) V. \tag{10}$$

Quantum particle swarm optimization (QPSO)

Particle swarm optimization (PSO) is a population-based stochastic optimization algorithm [55] whereby particle, representing a potential solution, which will be evaluated based on its fitness value and compared against both its individual best and the global best found by the entire swarm. This comparison guides the particles towards promising regions of the search space. However, reliance on both particle position and velocity can confine particles to a restricted area, particularly if their velocity remains static. This restriction can hinder the exploration of the entire solution space, potentially leading to local optima stagnation.

Quantum-inspired particle swarm optimization (QPSO) is a powerful computation technique that introduces quantum mechanics theory into particle swarm optimization (PSO), allows particles to explore the solution space with increased flexibility. Unlike the fixed trajectories of traditional PSO, particles in QPSO can exhibit uncertain movements, appearing in different locations within the search space. This effectively prevents them from getting trapped in local optima; hence, leading to improved global search capabilities [56–58]. In QPSO, a wave function is used to represent the motion state of particles. Since space and time are independent of each other in quantum space, particle corresponding to the wave function are also considered as random. In addition, QPSO has the advantages of having fewer parameters, a simple structure, and a faster convergence rate.

The QPSO algorithm introduces a parameter m_{best} to represent the average value of the best historical positions p_{best} of all particles in the swarm. The following steps outline the particle update process in the QPSO algorithm [59]:

Step 1: Calculate m_{best}, i.e,

$$m_{best} = \frac{1}{N} \sum_{i=1}^{N} p_{best_i}, \tag{11}$$

where N is the number of particles in the swarm; p_{best_i} represents the ith particle's personal best position in the current iteration.

Step 2: Update particle position u:

$$P_i = \phi * p_{best_i} + (1 - \phi)g_{best}, \tag{12}$$

where g_{best} refers to the current best particle in the entire swarm; P_i updates the position of the ith particle. The particle position update formula is

$$x_i = P_i + \alpha|m_{best} - x_i|ln\left(\frac{1}{\mu}\right), \tag{13}$$

where x_i denotes the position of the ith particle; is updated using the innovation parameter α; and two uniformly distributed random numbers ϕ and μ (both in the range (0,1)), the probability of these random numbers being positive or negative is equal to 0.5. It can be seen that there is only one innovative parameter setting α known as the contraction-expansion (CE) coefficient, which can be tuned to control the convergence speed of the algorithms, and α is generally not greater than 1.

Long short-term memory (LSTM) network structure with quantum particle swarm optimization

When using the QPSO algorithm for the parameter tuning, particle initialization is transformed into a series of LSTM parameters, and its fitness is the R2 score value of the LSTM model when using initialization parameters, i.e.,

$$\text{Fitness(QPSO)} = R^2 = \frac{SSR}{SST} = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}, \tag{14}$$

where $\hat{y}_i$ represents the predicted value, y_i represents the actual values, $\bar{y}$ represents the mean of all the values and N represents the number of training sets. QPSO-LSTM model has some hyperparameters to optimize, such as the time step *TS*, the number of hidden layer nodes L_1, L_2, batch size B as shown in Algorithm 1. QPSO can be used to quickly determine the hyperparameter combination suitable for the time prediction model, so as to effectively improve the accuracy of the prediction model. The flow chart for optimizing the LSTM model with the QPSO optimization algorithm is shown in Fig. 1.
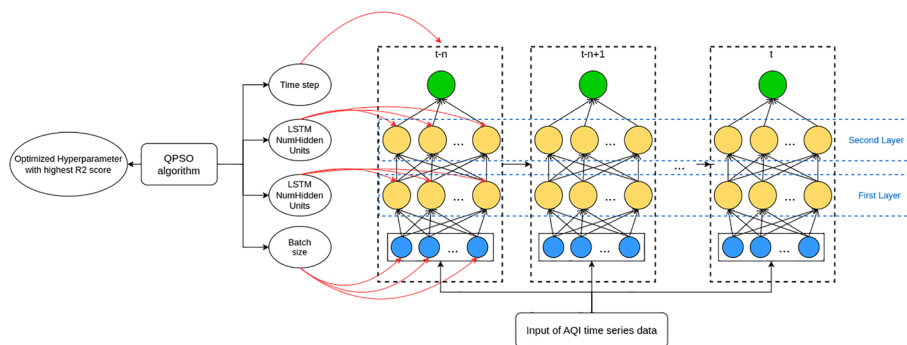
Fig. 1 The structure of the QPSO-LSTM model

Algorithm 1 Quantum particle swarm optimization for LSTM

Input:
  1: Time step TS
  2: Number of hidden layer nodes L_1, L_2
  3: Batch size B
Output:
  4: Optimized parameters $TS_{opt}, L_{1,opt}, L_{2,opt}, B_{opt}$ with the best R2 score
Initialize:
  5: The population size M
  6: The positions of the particle P_i
  7: The dimension of the particles D
  8: The contraction-expansion (CE) α
  9: Maximum iteration T
Begin
  10: **for** $t = 1$ to Maximum Iteration T **do**
  11:     Compute the mean best position C;
  12:     **for** $i = 1$ to population size M **do**
  13:        **if** $f(X_i) < f(P_i)$ then $P_i = X_i$ **then**
  14:        **end if**
  15:        $G = argmin(f(P_i))$
  16:        **for** $j = 1$ to D **do**
  17:           $\phi = rand(0,1); u = rand(0,1)$
  18:           $p_{ij} = \phi.P_{ij} + (1 - \phi).G_j$
  19:           **if** $(rand(0,1) > 0.5$ **then**
  20:              $X_{ij} = p_{ij} + \alpha.|C_j - X_{ij}|.log(1/\mu)$
  21:              $X_{ij} = p_{ij} - \alpha.|C_j - X_{ij}|.log(1/\mu)$
  22:           **end if**
  23:        **end for**   ▷ end for loop j
  24:     **end for**   ▷ end for loop i
  25: **end for**   ▷ end for loop t
End

Model integration development

Consider the time series data x_t as the combination of linear component L_t and nonlinear component N_t represented by mathematical (15).

$$x_t = L_t + N_t. \tag{15}$$

Though linear and non-linear modeling methods specialize in different types of patterns, their combined application can yield a comprehensive understanding of the intricate features within a time series [60]. The ARIMA model can predict short-period linear trends well, while the LSTM model can predict complex, non-linear time series well, c.f. [61]. The ARIMA model is used to predict the linear and nonlinear components of the data, which is then fed into the deep neural network and fit to obtain the predicted value of the nonlinear component. On this basis, both linear and nonlinear aspects of data are integrated, and the final prediction result is obtained. To overcome the blindness of hyperparameter setting, the QPSO algorithm is introduced in this research to determine the optimal value of hyperparameter setting, the model flow is shown in Fig. 2.

The AQI prediction architecture utilizes a pre-trained ACNN-LSTM model based on a sequence-to-sequence framework. The ACNN encoder extracts deep features via
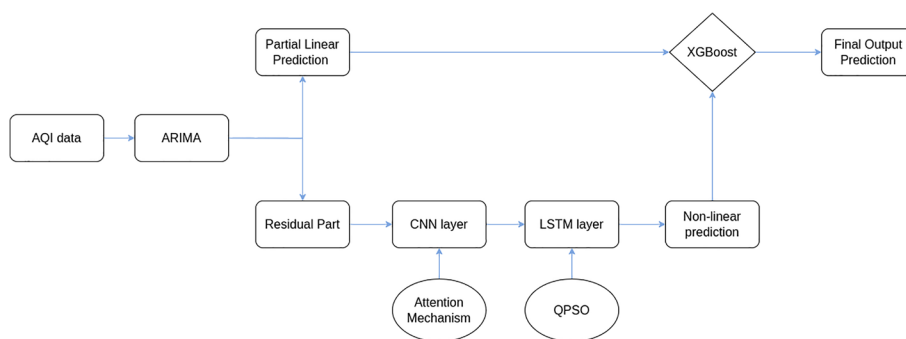
Fig. 2 ARIMA-ACNN-QPSO-LSTM-XGBoost model and derived model process

convolutional layers, while the bidirectional LSTM decoder learns long-term temporal dependencies. The encoder–decoder architecture mitigates noise, and the deep learning approach effectively captures hidden state information, despite not fulfilling the air quality's linear property assumptions. Notably, the LSTM decoder receives context information from the ACNN encoder. The ACNN encoder utilizes a self-attention layer and CNN to compute context vectors (Q, K, V) and hidden states (H), as detailed in Equations (7) and (10). Combined with previous decoder outputs, these drove the LSTM decoder's predictions. Multi-head attention within the encoder captures the relationships between current and past sequences and embeddings, while masked attention during decoding restricts the decoder's view to previously processed elements. The key insight behind our ACNN encoder is its ability to overcome LSTM limitations, c.f. [62]. ACNN, equipped with multi-head self-attention and multi-scale convolutions, excels at capturing local and global dependencies, mimicking the human cognitive system's focus on salient features. LSTM, on the other hand, handles temporal dynamics effectively. This synergy enhances both structural and time-series modeling capabilities. After decoding, a XGBoost regressor, known for its flexibility and strong learning power, further extracts hidden features and fine-tunes the model, leading to superior prediction accuracy and generalizability for air quality data [63].

Data and implementation

Study area and data

The focus of this research is on Seoul's urban region, where air pollution data from 2021 to 2022 were collected from Seoul Air Data [64], which is a centralized repository established in 2021. The data set comprises the hourly concentrations of six components of air quality O_3, CO, NO_2, SO_2, PM_{10} and $PM_{2.5}$. Data were collected from 25 air pollution monitoring stations in Seoul, representing the 25 district stations in the city, as depicted in Fig. 3. Additionally, the data set also contained meta-information on the data including the station location and timestamp of the concentrations that were recorded. Before using the data for analysis and modeling, we had assessed their quality and reliability.

The distribution of air quality monitoring stations can influence the reported pollutant concentrations. Stations situated in densely populated areas tend to record higher pollutant levels compared to those located in areas with more greenery or preserved natural spaces. The accompanying map illustrates the station locations
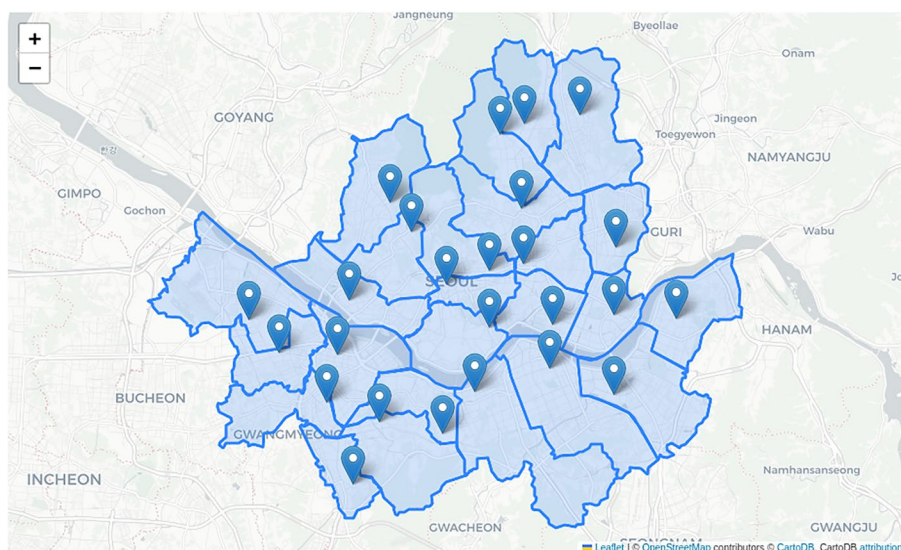
Fig. 3 Air quality monitoring stations in Seoul

superimposed on district borders. As evidenced from the map, station placement within each district is uneven, resulting in varying degrees of spatial coverage. For example, Dongdaemungu and Jongnogu stations in Seoul's central region are positioned close to each other, while a substantial area in the south remains uncovered. To avoid underestimating the potential severity of air quality issues, we will adopt the maximum AQI value across all 25 stations as the representative metric for our analysis. This approach accounts for variability across locations and mitigates the risk of overlooking localized pollution events.

Exploratory data analysis

In this research, exploratory data analysis (EDA) was conducted to provide insights into data characteristics and spatiotemporal patterns in the air quality data. Figures 4 and 5 illustrates the overall trends in O_3, CO, NO_2, SO_2, PM_{10}, and $PM_{2.5}$ in the specified research periods. The amount of air pollution is reduced from August 2021 to November 2021 mainly due to the Covid-19 shutdown in the city. Despite having good overall pollutant levels, the plot reveals that Seoul can occasionally experience poor or very poor concentrations of gases and particles.

It is well established that automobiles and industrial activities are major contributors to air pollution, often emitting multiple pollutants simultaneously. For instance, motor vehicles typically release both CO and NO_2. Considering this shared source of emissions, we anticipate a degree of correlation between the time series of these pollutants.

The correlation matrix in Fig. 6 reveals that most pollutant pairs exhibit absolute correlations exceeding 0.3, indicating a strong positive relationship between them. The most prominent correlations are observed between CO and NO_2, as well as PM_{10} and $PM_{2.5}$. These findings corroborate our assumption that vehicular emissions significantly contribute to urban air pollution.
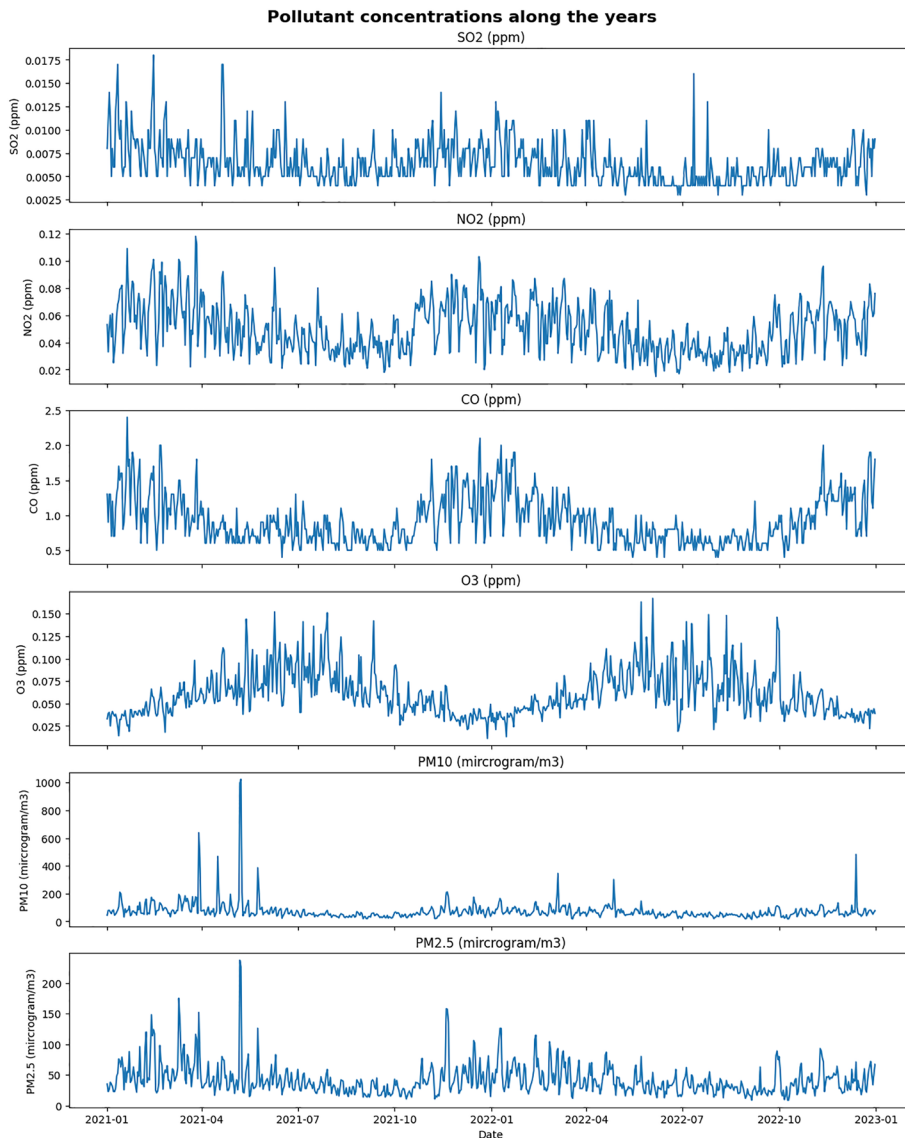
Fig. 4 Daily air quality in Seoul (from 2021 to 2022)

The limited duration of available data made the monthly forecasting ineffective in enhancing model performance. Consequently, hourly data was incorporated as a characteristic in the construction of the model for each pollutant.

Figure 7 shows the distribution of pollutant concentrations observed over a 2-year period in Seoul. While the overall air quality appears satisfactory, the plot reveals instances where pollutant levels reach unhealthy or very unhealthy categories, especially $PM_{2.5}$ and PM_{10}. To facilitate spatial analysis of the data, the mean concentration of each pollutant was determined for each region. However, a challenge arose due to the disparate units of measurement employed for the various pollutants. To address this inconsistency and enable meaningful comparisons, the six pollutant distributions, each encompassing data for 25 districts, were standardized. In Fig. 8, pollution levels are expressed in standardized units representing deviations from the mean. These units, known as z-scores [65], quantify the relative position of
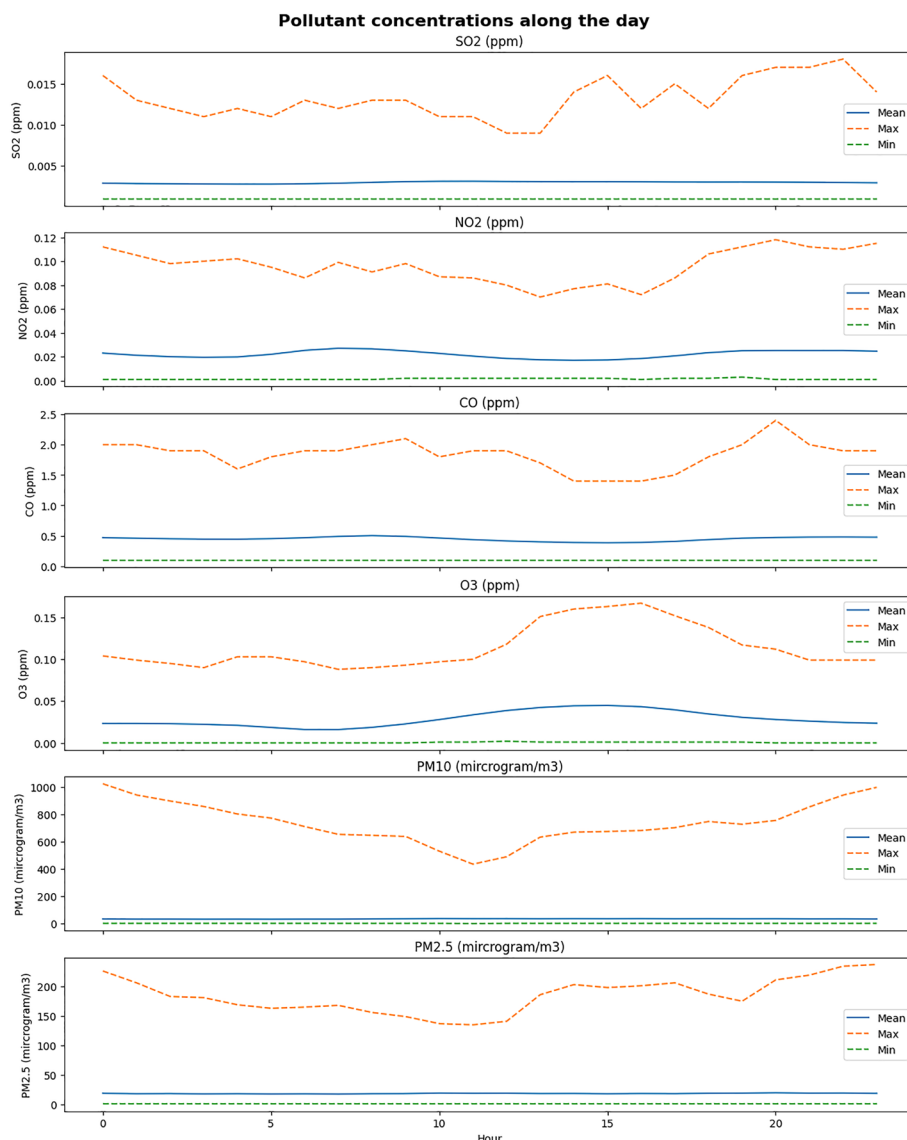
Fig. 5 Hourly air quality in Seoul (from 2021 to 2022)

a specific pollution measurement within the overall distribution. Negative z-scores indicate pollution levels lower than the average, while positive z-scores correspond to levels exceeding the mean. Consequently, based on the aforementioned considerations regarding location and PM2.5 pollution risk, Nowon-gu and Dongjak-gu were chosen for comparative analysis of station AQI prediction. These districts are situated in the northern and southern regions of South Korea, respectively.

Data preprocessing

The quality of time series data is crucial to develop accurate time series forecasting models. It directly impacts model performance and the reliability of parameter estimations. In the context of air quality data, outliers, unstructured timestamps, and missing values pose significant challenges. These issues can arise from various
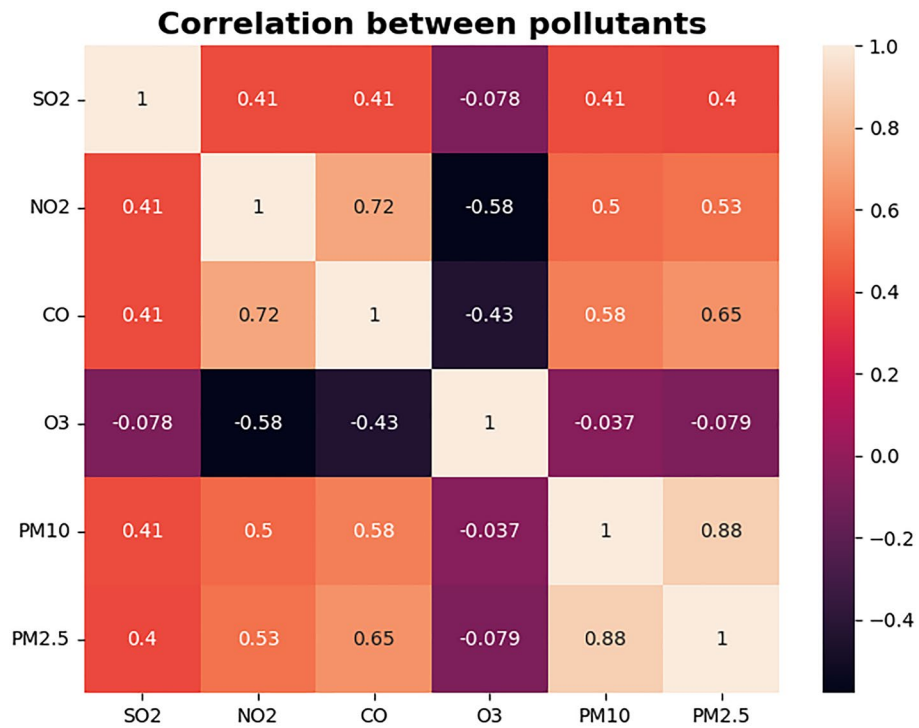
Fig. 6 Correlation between pollutants in Seoul
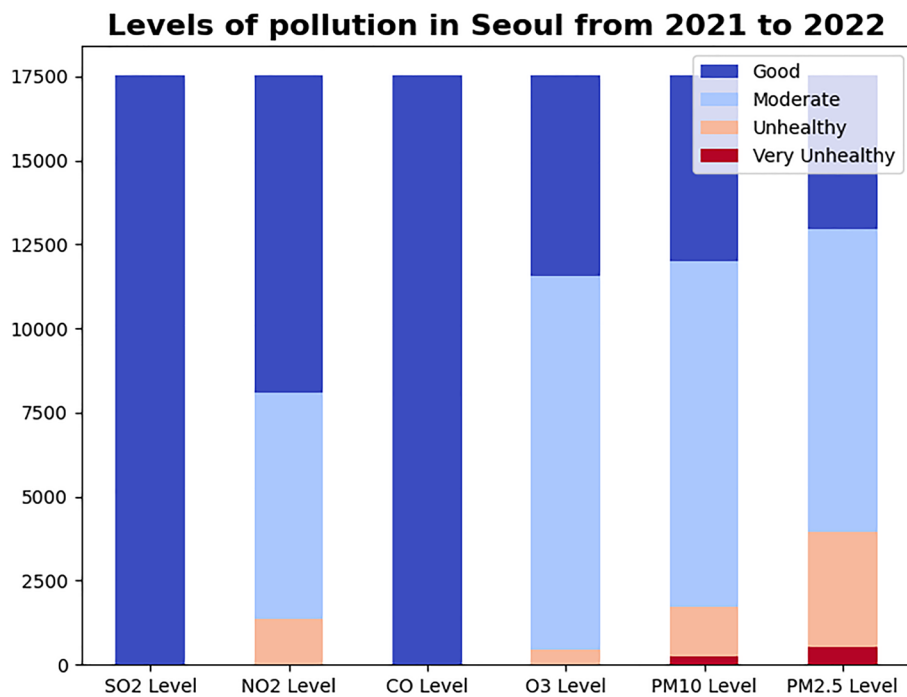
Fig. 7 Pollutant levels in Seoul from 2021 to 2022
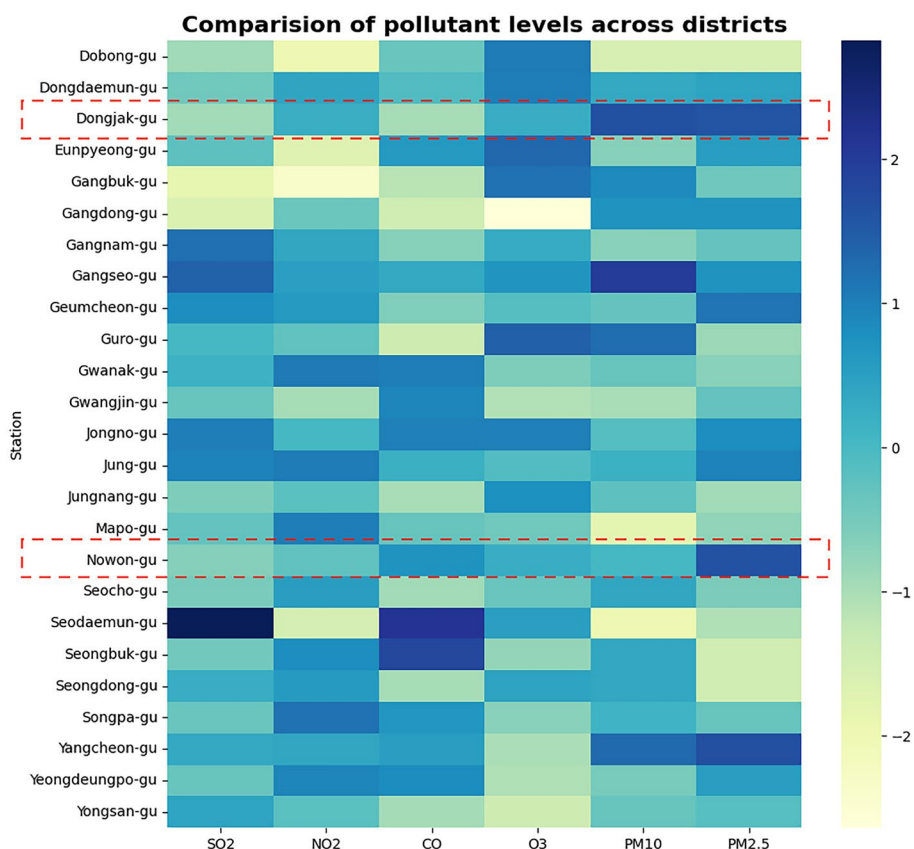
Fig. 8 Pollutant levels comparison in 25 stations

factors, including monitoring station malfunctions or external influences, leading to inaccurate or incomplete air quality measurements. When creating a hybrid deep learning model, the following pre-processing procedures were taken:

- Timestamps play a crucial role in time series modeling, as they provide temporal context for the data. To ensure compatibility with subsequent data processing and analysis, the raw timestamp data were first converted into a standardized date-time format.
- All negative values were eliminated and replaced with 'nan,' which were treated as missing values.
- The presence of outliers can be attributed to various factors, such as unexpected events or technical glitches, resulting in abrupt fluctuations in the trend line that significantly deviate from the overall data pattern. To address this issue, these outlier values were replaced by 'nan,' effectively classifying them as missing data.
- The existence of missing values may influence prediction performance. Therefore, appropriate missing data imputation techniques were applied to the training dataset before model development. For isolated missing values, the most recent available value was imputed. In cases where multiple consecutive hours lacked data, the corresponding values of the same hours on the previous day were used for imputation.

The data used in this article comes from the open and free public dataset [64] for the research of air quality in Korea, which has the characteristics of rich data, simple use and convenient implementation. The data is selected from the data from January 1st 2021 to December 31st 2022, the data in 1 h denotes a point of the sequence. The train set and test set was divided by 80–20, which means the training set comprises data collected from January 1st 2021 to August 6th 2022 and testing set consists of the remaining data from August 7th 2022, onwards. The data undergoes normalization to a range of (0, 1] using the following formula:

$$X_{\text{scaled}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}}. \tag{16}$$

After model training and prediction, an inverse normalization step is required. This allows for the calculation of evaluation functions and the plotting of results. The inverse normalization equation is:

$$X_{\text{pred}} = X_{\text{scaled}} * (X_{\max} - X_{\min}) + X_{\min}. \tag{17}$$

To evaluate the performance between stations, AQI calculation will be provided. The AQI scale typically ranges from 0 to 500, where lower values represent good air quality and higher values indicates poorer air quality. The scale is divided into categories that indicate different levels of health concern, which is shown in Table 1 . The AQI for each pollutant is calculated by interpolating the measured concentration between the breakpoints that define the AQI categories. The formula for interpolation is: [7]

$$I_p = \frac{(I_{HI} - I_{LO}) \times (C_p - BP_{LO})}{BP_{HI} - BP_{LO}} + I_{LO}, \tag{18}$$

where I_p is the air quality index for each target pollutant, BP_{HI} and BP_{LO} are the AQI values corresponding to the higher and lower concentration breakpoints for the pollutant's AQI categories. C_p is the rounded pollutant concentration. I_{HI} and I_{LO} are the concentration breakpoints that C_p falls between.

Performance measure metrics

In this article four common performance measures are presented: mean absolute error, mean squared error, root mean squared error, R2 score to evaluate the accuracy of the different algorithms. Given the n set of predictions $y_1, \ldots, y_n$ made by a model, we can define the following formula [66]:

Mean Absolute Error:

$$\text{MAE} = \frac{\sum_{i=1}^{n} |y_i - \hat{y}_i|}{n}. \tag{19}$$

Mean Squared Error:

Nguyen *et al. Journal of Big Data*   (2024) 11:71

Page 20 of 38

Table 2 Default model hyperparameter settings

Model type	Models	Hyperparameter setting
Deep learning model	LSTM	Neuron1 = 50, neuron2 = 50, batch size = 32. epochs=30 with EarlyStopping, learning rate = 0.001, sliding window = 10
	BiLSTM	neuron = 50, batch size = 32. epochs = 50 with EarlyStopping, learning rate = 0.001, sliding window = 10
	CNN-LSTM	Conv1D filtes=64, kernel size = 1; MaxPooling1D pool size = 2; LSTM neuron = 50; batch size = 32, epochs = 50 with EarlyStopping, learning rate = 0.001, sliding window = 10
Optimization algorithm	QPSO	number of particle = 30, dimension of particles = 4, alpha = 0.55, max iteration = 5, lstm iteration number = 30 with EarlyStopping, sliding window=[1,50], neuron1 = [1,300], neuron2 = [1,300], batch size = [1,512]

Table 3 Statistical ARIMA model parameter settings

Factors	SO_2	NO_2	CO	O_3	PM_{10}	$PM_{2.5}$	Nowon-gu	Dongjak-gu
Settings	(2,0,3)	(2,0,3)	(5,0,0)	(2,0,1)	(1,0,4)	(1,0,5)	(2,0,2)	(2,0,2)

$$\text{MSE} = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}. \tag{20}$$

Coefficient of determination:

$$\text{R}^2 = 1 - \frac{\text{RSS}}{\text{TSS}} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y_i})^2}, \tag{21}$$

where y_i is the actual value, $\hat{y}_i$ is the predicted value, $\overline{y_i}$ is the mean value of a sample, n is the number of observations, RSS is the residual sum of squares, TSS is the total sum of squares.

Experiment setup

The hybrid DL model was implemented in Python on a robust computing platform, which includes NVIDIA RTX 3080 GPU, Intel Core i9-13900K CPU, to ensure accurate and efficient forecasting.

The default model hyperparameter settings were selected and described in Table 2. We have used ADAM [67] as optimizer with default momentum as presented in the paper.

Results and discussion

To demonstrate the benefits of the ARIMA-ACNN-QPSO-LSTM model, we compared its performance against classic machine learning, deep learning, and statistical models. We employed a one-dimensional regression equation and conducted multiple experiments for each model to optimize their forecast accuracy. For the statistical model, we used the AIC criterion to restrict the maximum values of p, q to 5 and d to 2, ensuring a fair comparison across all models. Table 3 shows all the optimal parameters for ARIMA model in each AQI criteria. After performing an ARIMA model fitting on the historical

data, several potential models were obtained. AIC was employed to evaluate the models, and the model with the minimum AIC value was selected.

Both SO_2 and NO_2 have the same ARIMA settings of (2,0,3). This indicates a moderate level of autocorrelation where the current value is significantly influenced by the immediate two past values. The absence of differencing (d = 0) suggests that the series for these pollutants is stationary, requiring no differencing to achieve stationarity. The moving average component (q = 3) indicates that the prediction error is influenced by the error terms of the three previous forecasts. This similarity could suggest that SO_2 and NO_2 share similar emission sources or atmospheric behaviors. The ARIMA(5,0,0) of CO model suggests a strong linear dependency on the previous five values, with no need for differencing or moving average components. With an ARIMA(2,0,1) setting, O_3 shows a reliance on the immediate past values and a slight adjustment based on the error of the previous forecast. PM_{10} and $PM_{2.5}$ show a more significant dependence on the moving average component. For Nowon-gu and Dongjak-gu station, the uniform ARIMA(2,0,2) settings for these locations suggest similar air quality behavior in terms of temporal dynamics.

Next, we presents a comparative analysis of the proposed model's performance metrics against those of established statistical and DL models. Evaluation constitutes a

Table 4 Forecasting results obtained using the air pollution data from Seoul

Pollutant and station	Metrics	ARIMA	XGBoost	LSTM	BiLSTM	CNN-LSTM	QPSO-LSTM	QPSO-BiLSTM	Proposed model
SO_2	MSE	4.36e−7	4.3e−7	4.29e−7	4.28e−7	4.66e−7	4.37e−7	4.21e−7	4.20e−7
	MAE	4.41e−4	4.59e−4	4.49e−4	4.50e−4	4.80e−4	4.34e−4	4.23e−4	4.09e−4
	R^2	0.65641	0.51472	0.51981	0.52250	0.40533	0.56061	0.62351	0.66544
NO_2	MSE	2.01e−5	1.66e−5	1.82e−5	1.76e−5	2.47e−5	1.80e−5	1.73e−5	1.56e−5
	MAE	0.00317	0.00299	0.00309	0.00302	0.00362	0.00309	0.00302	0.00291
	R^2	0.92025	0.93330	0.92722	0.92775	0.90134	0.92736	0.92941	0.93737
CO	MSE	0.00865	0.00686	0.00799	0.00791	0.01011	0.00759	0.00743	0.00681
	MAE	0.05898	0.05590	0.06254	0.06108	0.07160	0.05988	0.05954	0.05583
	R^2	0.87282	0.88910	0.87190	0.87730	0.83430	0.88503	0.88601	0.89980
O_3	MSE	2.18e−5	2.16e−5	2.37e−5	2.26e−5	3.26e−5	2.16e−5	2.09e−5	2.15e−5
	MAE	0.00340	0.00332	0.00349	0.00337	0.00417	0.00331	0.00328	0.00333
	R^2	0.92561	0.92930	0.92032	0.92201	0.88183	0.92393	0.92664	0.93000
PM_{10}	MSE	48.61285	39.99597	68.03935	60.55632	72.97267	60.86984	52.43590	39.82543
	MAE	4.26909	4.65191	5.20008	4.83833	5.35122	5.09820	4.52110	3.66593
	R^2	0.93943	0.94017	0.90935	0.92988	0.89975	0.92298	0.93106	0.95038
$PM_{2.5}$	MSE	13.24212	9.27569	13.39572	11.82768	17.09323	12.70678	11.05409	9.01954
	MAE	2.53490	2.39400	2.59168	2.39453	2.86955	2.46828	2.33176	2.17691
	R^2	0.93825	0.94721	0.93354	0.94252	0.91977	0.94217	0.94449	0.95794
Nowon-gu	MSE	103.94584	104.95484	129.70048	129.09137	214.36079	122.99369	108.03047	6.78195
	MAE	6.57444	6.67554	8.11201	7.99737	9.98624	7.87624	7.34626	1.24056
	R^2	0.90578	0.89952	0.87176	0.85430	0.79713	0.88132	0.90079	0.99385
Dongjak-gu	MSE	89.42425	89.92425	202.26346	140.64929	273.84135	147.56597	129.73990	6.09202
	MAE	5.86810	6.16812	9.85402	7.74239	11.50207	8.32623	7.69737	1.23124
	R^2	0.92753	0.92397	0.75669	0.83737	0.70835	0.87518	0.87649	0.99506

(a) MSE of SO2 prediction   (b) MAE of SO2 prediction   (c) R-squared of SO2 prediction
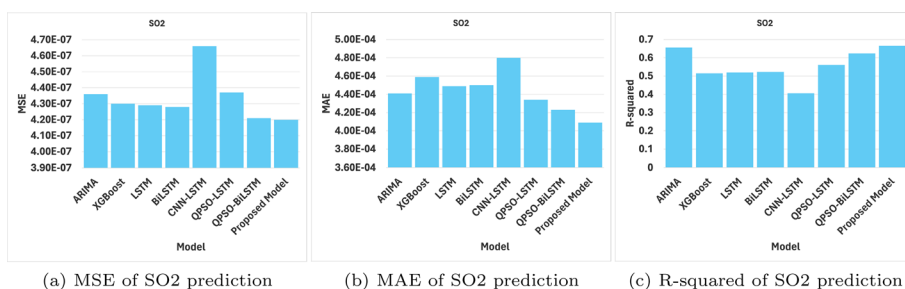
Fig. 9 Visualize SO$_2$ forecast metrics comparison

(a) MSE of NO2 prediction   (b) MAE of NO2 prediction   (c) R-squared of NO2 prediction
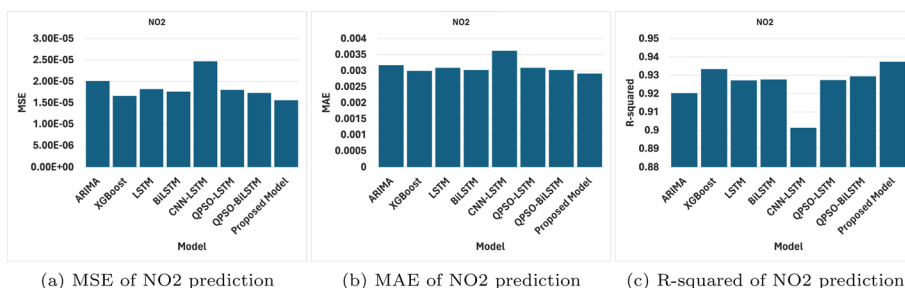
Fig. 10 Visualize NO$_2$ forecast metrics comparison

(a) MSE of CO prediction   (b) MAE of CO prediction   (c) R-squared of CO prediction
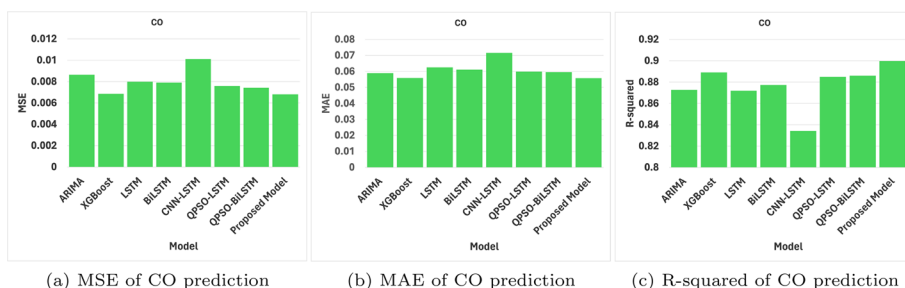
Fig. 11 Visualize CO forecast metrics comparison

(a) MSE of O3 prediction   (b) MAE of O3 prediction   (c) R-squared of O3 prediction
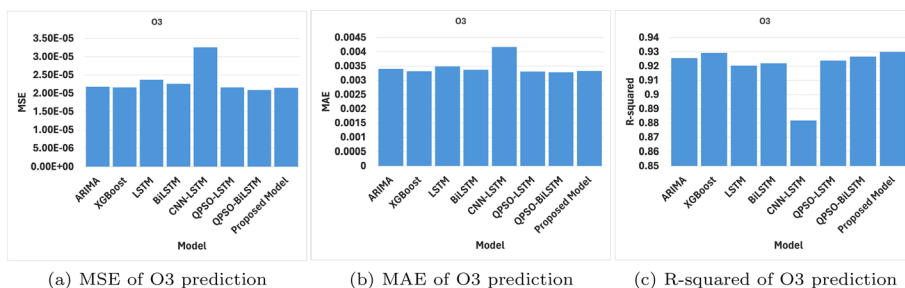
Fig. 12 Visualize O$_3$ forecast metrics comparison

fundamental step in model implementation, as it enables the identification of the optimal model based on its demonstrated capabilities. Model performance was evaluated by comparing actual values with predicted outcomes. To facilitate this comparison,

(a) MSE of PM10 prediction   (b) MAE of PM10 prediction   (c) R-squared of PM10 prediction
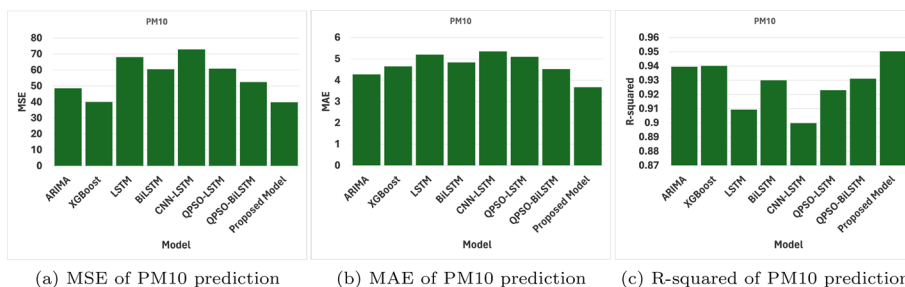
Fig. 13 Visualize PM_{10} forecast metrics comparison

(a) MSE of PM2.5 prediction   (b) MAE of PM2.5 prediction   (c) R-squared of PM2.5 prediction
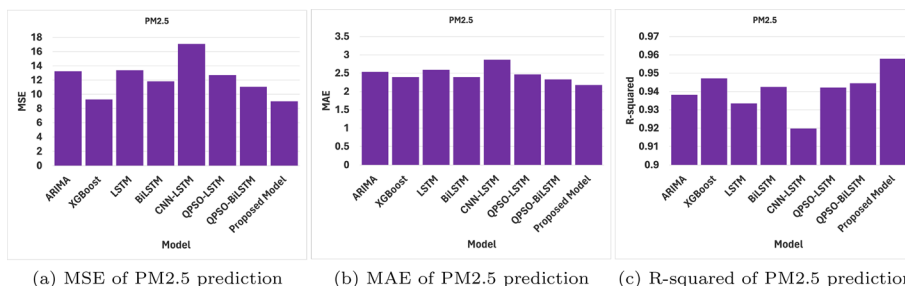
Fig. 14 Visualize $PM_{2.5}$ forecast metrics comparison

(a) MSE of AQI prediction   (b) MAE of AQI prediction   (c) R-squared of AQI prediction
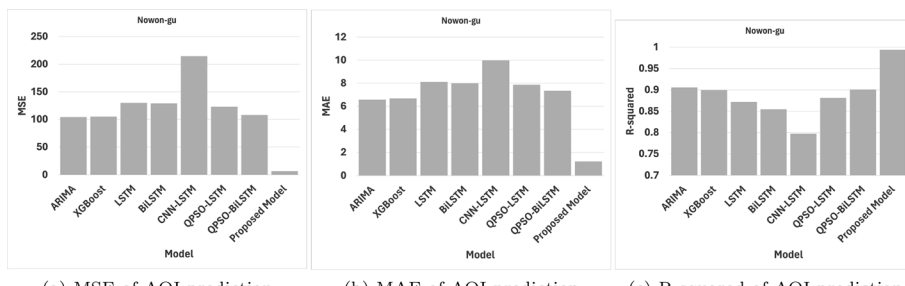
Fig. 15 Visualize AQI Nowon-gu station forecast metrics comparison

(a) MSE of AQI prediction   (b) MAE of AQI prediction   (c) R-squared of AQI prediction
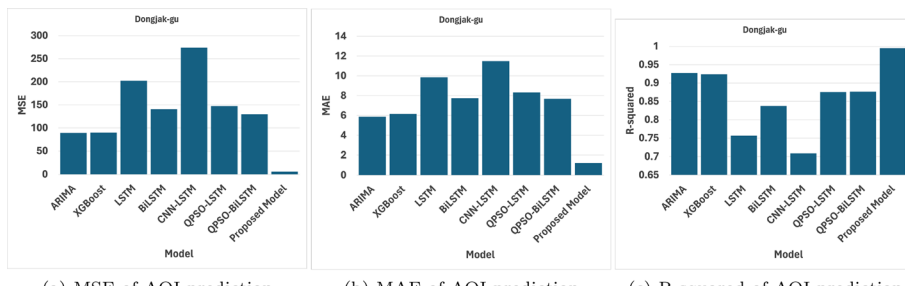
Fig. 16 Visualize AQI Dongjak-gu station forecast metrics comparison

three crucial metrics are employed in the analysis. The optimal result that are obtained after carefully evaluating with various AQI concentration and comparison of the scores of the models as shown in Table 4. The results are further visualized in a bar chart, as

shown in Figures 9, 10, 11, 12, 13, 14, 15 and 16. Our findings show that the MAE, MSE and R^2 of our proposed model is lower than all other models. Hence, it is reasonable to assume that the proposed model is appropriate for projecting data related to air quality. The model consistently maintains low error rates (e.g., MSE, MAE) and high coefficient of determination (R^2) values across all AQI pollutants, signifying its superior predictive power. Even with minimal data, the model's predictions remain remarkably accurate, indicating its suitability for scenarios with limited data availability.

We compared the evaluation indicators generated by the evaluation function with the output results of all models after their execution. In order to clearly compare the hybrid

(a) Comparison of different AQI models

(b) ARIMA

(c) LSTM

(d) CNN-LSTM

(e) BiLSTM

(f) QPSO-LSTM

(g) QPSO-BiLSTM

(h) XGBoost

(i) Proposed Model
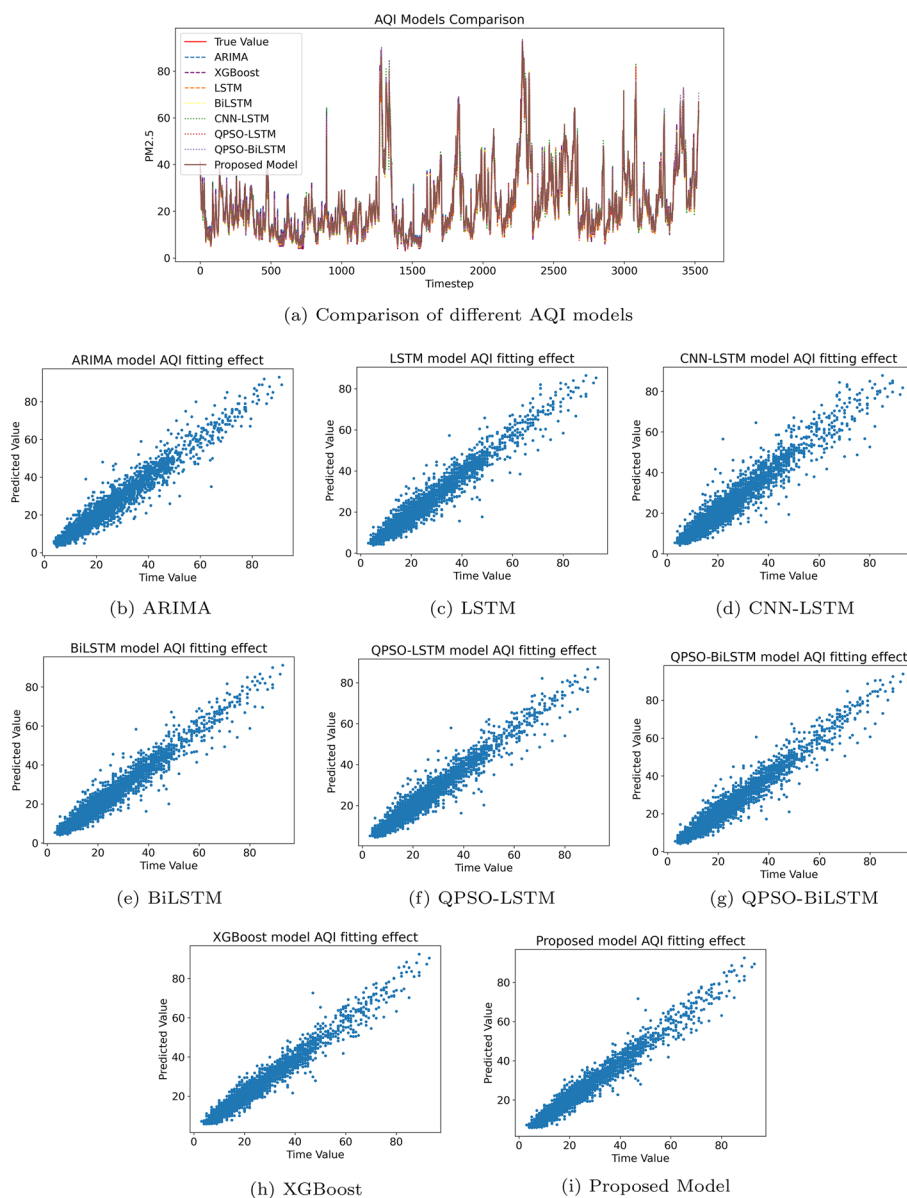
Fig. 17 PM$_{2.5}$ concentration forecast comparison

model with other models, the best-performing model among the models was selected to plot line and scatter plots, as shown in Fig. 17 and Appendix.

As shown in Table 4, compared to the single model's limited forecasting ability, the combined model delivers significantly more accurate predictions. For each pollutant, we take the best values, which are lowest for MSE, MAE and highest for R^2, from the other models and compare it with the value from the proposed model. For $PM_{2.5}$ the hybrid proposed model has a 31.13% reduction in MSE, 10% reduction in MAE and 1.64% improvement in R^2 relative to the best DL model (Bi-LSTM). The QPSO can improve the prediction accuracy of the model, for example for LSTM model, the MSE metric is reduced by 5.42%, MAE is reduced by 5.28% and R^2 is improved by 1%. As can be seen in Fig. 17, our hybrid model excels at predicting $PM_{2.5}$ compared to both the single and combined models, and its scatter plot shows the clearest data clustering. For SO_2 prediction, the proposed model outperforms other best models with the lowest MSE and MAE with 0.5% and 3.31% improvement, and a competitive R^2 value, indicating higher accuracy and reliability even with the very small value point. For NO_2 prediction, all models perform comparably well, with R^2 values all above 0.92. The proposed model again shows a advantage in 9.83% improvement in MSE, 3.64% improvement in MAE and slight increase 0.9% in R^2, suggesting its potential for more precise predictions. We have identified the narrow range of SO_2 values in dataset, which seen in Figure 21, (from 0.000 to 0.015) as a significant factor contributing to not predict values well. This limited range presents a unique challenge for the model. Given the small magnitude of change within SO_2 concentrations, the model must achieve a high level of precision to accurately predict these values. This is a more strenuous task compared to pollutants with a broader range of values, where slight inaccuracies in prediction could not significantly impact the overall performance metrics like R-squared. O_3 forecasts indicate a tight competition among models, with small or no improvement in term of all evaluated metrics. In the case of CO, our proposed model witness an considerably improvement of 12% in MSE and 5.1% in MAE. With the QPSO-enhanced, models demonstrate slightly better performance in minimizing errors, as evidence by their high R^2 value compared to default models. For PM_{10} prediction, all models have well performance with very high R^2 value. Also, our proposed model shows significant improvement of 19.03% in MSE and MAE, and slight improvement of R^2 of 2%. In summary, the proposed model generally exhibits superior performance across all pollutants, with consistently lower error rates and high R^2 values. This indicates its robustness and reliability as a forecasting tool. The QPSO-enhanced models also show strong performance, particularly in explaining the variance of data. These findings underscore the potential of advanced computational models in environmental monitoring and the importance of selecting appropriate models for different pollutants to achieve the best forecasting results.

As discussed in "Quantum particle swarm optimization (QPSO)" section, QPSO overcomes the problem of premature convergence in local optima, a common issue with traditional PSO. Its probabilistic movement mechanism allows particles to break free from suboptimal regions and explore diverse areas of the search space, significantly increasing the chances of finding the global optimum without requiring additional networks [41]. The proposed model shown in Fig. 2 leverages the strengths of two distinct models, each specializing in different aspects of air quality prediction. This

combination, using re-average XGBoost model to solve for individual limitations [47], ultimately shows more accurate and comprehensive forecasts.

For Nowon-gu station, the proposed model significantly outperforms the others with the lowest MSE (6.78195) and MAE (1.24056), indicating it predicts the air pollution levels with the least error. Traditional models like ARIMA and XGBoost, along with deep learning models such as LSTM and BiLSTM, show much higher errors. Particularly, CNN-LSTM has the highest MSE (214.36079), suggesting it is the least effective model for this dataset. The proposed model also achieves an R-squared score of 0.99385, nearing perfect predictive ability. Other models show considerable variability in performance, with ARIMA and QPSO-BiLSTM being the potential models. Similar to Nowon-gu, for Dongjak-gu station, the proposed model dramatically surpasses other models in accuracy for Dongjak-gu, with the lowest MSE (6.09202) and MAE (1.23124). It is noteworthy that the LSTM-based models and their enhancements (like QPSO-LSTM) generally performed poorly in terms of MSE, particularly CNN-LSTM, which had the highest MSE (273.84135). The proposed model again leads with an R-squared score of 0.99506, indicating good prediction accuracy. This is a significant improvement over the other models, where the performance again varies, with ARIMA and QPSO-enhanced models showing relatively better but still substantially less effective results than the proposed model.

The proposed model demonstrates a substantial improvement over traditional statistical models (like ARIMA) and various machine learning models, including those based on LSTM and its variants. This indicates that the proposed model might incorporate a more sophisticated mechanism for capturing and forecasting air pollution levels, potentially accounting for nonlinearities and complexities in the data that other models fail to address effectively. The high R-squared values achieved by the proposed model for both stations suggest that it can explain a vast majority of the variance in air pollution levels, making it a potentially valuable tool for environmental monitoring and policy-making. The significant disparity in performance between the proposed model and other models, especially in terms of MSE and MAE, underlines the importance of model selection and the potential for innovative approaches to provide breakthroughs in environmental data analysis. It is also worth noting that while LSTM and its variants are generally considered powerful for time series forecasting due to their ability to capture temporal dependencies, their performance in this instance was outclassed by the proposed model. This could be due to specific features or architectures of the proposed model that are particularly well-suited to forecasting air pollution data.

In summary, the results underscore the efficacy of the proposed model in forecasting air pollution levels with high accuracy, suggesting its value for practical applications in environmental science and public health. Further research could explore the specific features and methodologies of the proposed model that contribute to its superior performance, as well as its applicability to other types of environmental data and forecasting challenges.

Conclusion

Air pollution is a persistent environmental challenge across the globe and accurate Air Quality Index (AQI) predictions has a crucial role in effective air pollution management. Precise and consistent AQI predictions are vital not only for public health in our cities,

but also for ensuring the environment's long-term resilience against air pollution's detrimental impacts. While conventional time series models for air quality forecasting often have large prediction error, emerging neural networks represented by Long Short Term Memory (LSTM) have revolutionized the field with impressive accuracy. In this paper, we propose a new model based on the advantage of statistical method, deep learning and machine learning to predict the AQI concentration of different pollutants in Seoul, South Korea. Our objective was to conclusively demonstrate the strengths of our model, instead of limiting the comparison to derivative models, we included a diverse range of benchmark models, including the most popular algorithms used in practice for air quality forecasting. We address the ARIMA model's limited ability to capture nonlinearities, achieving high accuracy while simultaneously reducing computational workload. Our innovative approach overcomes the traditional hurdles of noisy and short time series data, allowing neural networks to overpass in predicting even these challenging sequences.

Experimental results demonstrate the effectiveness of our method. We used the AQI data, which includes O_3, CO, NO_2, SO_2, PM_{10}, and $PM_{2.5}$ detected in Seoul, to construct and analyze all models and the experimental results show that our hybrid model has good prediction effect on the test set. The results show that our proposed model in this paper outperforms the comparison models in term of different evaluation metrics such as Mean Squared Error (MSE), mean absolute error (MAE) and coefficient of determination (R^2). For our proposed model, the MSE values of the AQI pollutants concentration are 4.20e−7, 1.56e−5, 0.00681, 2.15e−5, 39.82543 and 9.01954; the MAE values are 4.09e−4, 0.00291, 0.05583, 0.00333, 3.66593 and 2.17691; the R^2 values are 0.66544, 0.93737, 0.89980, 0.93000, 0.95038 and 0.95794. This proposed model predicts AQI levels with the best accuracy of any model and it can handle different types of air pollution situations.

In addressing the practicality of our proposed model, it is essential to highlight its direct applicability, efficiency, and adaptability within real-world settings. We have evaluated the operational feasibility of the model, focusing on its integration into existing South Korea air quality monitoring systems. The model's design allows for seamless deployment across various geographic locales and pollution scenarios, requiring minimal adjustments to accommodate local data characteristics. Emphasizing the model's scalability, we illustrate how it can support urban planning and public health initiatives by providing highly accurate, timely forecasts that enable responses to air pollution. By presenting case studies in South Korea and various application scenarios, we aim to demonstrate the and positive impact of our model, making a compelling case for its practicality in enhancing environmental monitoring and policy-making efforts. This study not only bridges the gap between theoretical innovation and practical application but also sets the stage for future advancements in the field, encouraging further exploration and adoption of our proposed model preventing air pollution.

Our research also has some limitations. Firstly, it is model generalization. While the proposed model shows superior performance on the dataset from Seoul, it is essential to test it across different geographic locations, pollution types, and temporal scales to assess its generalizability. Secondly, air pollution dynamics are influenced by a multitude of factors, including meteorological conditions, urban infrastructure, traffic patterns,

and industrial activities. The indicative results suggest that the proposed model captures these complexities well for the specific cases studied, but further research should investigate its adaptability to changing conditions and unforeseen events. Finally, the accuracy of forecasting models is heavily dependent on the quality and granularity of the input data. Additional datasets, particularly those with higher temporal resolution or more comprehensive environmental variables, could provide further insights into the model's performance and limitations.

The proposed model in this study has the following challenges and future work:

- Capturing long-range dependencies within the data can be challenging, potentially limiting the model's predictive power. Future work will focus on extending the proposed model to handle long-sequence data, enabling its application to tasks requiring analysis of temporal patterns and dependencies.
- To optimize our model's performance, future work could investigate on an exploration of various machine learning algorithms, seeking the best fit for the data and specific prediction task.
- Our analysis acknowledges the limitations inherent in excluding external factors such as meteorological indicators and seasonality from the model. Future work could incorporate these elements for a more comprehensive understanding of AQI fluctuations.
- Future studies could explore modifications to the proposed model or the development of hybrid models that combine the strengths of various approaches. Comparative studies involving additional datasets and alternative modeling techniques could expose more robust conclusions about the proposed model's efficacy.

In conclusion, while this study demonstrates the effectiveness of our model in outperforming traditional methods like BiLSTM with optimization fine-tuning, further exploration is needed. Integrating additional influencing factors, such as weather conditions and seasonality, can potentially achieve even greater accuracy and broaden the model's applicability. By acknowledging that our findings are indicative, we search for further scientific inquiry and collaboration. This stance encourages a proactive approach to model validation, the exploration of new data sources and modeling techniques, and the thoughtful consideration of the broader implications of our work on society and the environment. Continuous improvement and validation will be essential for advancing our understanding and developing effective tools for managing air pollution and its impacts.

Appendix: AQI forecast comparison
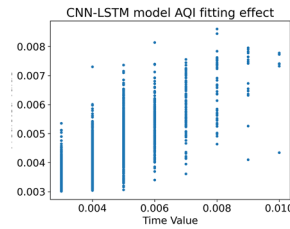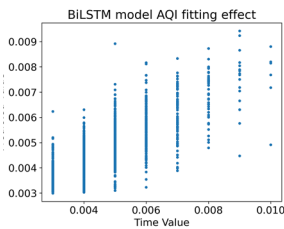
See Figs. 18, 19, 20, 21, 22, 23, 24.
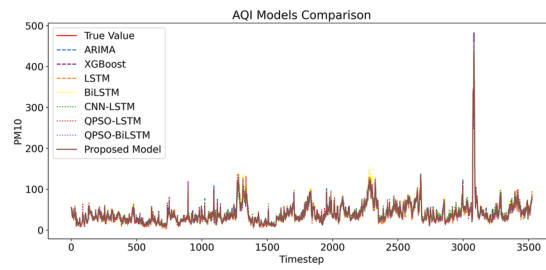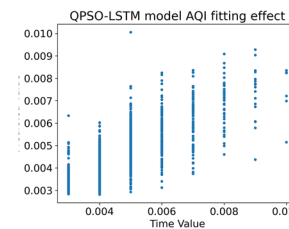
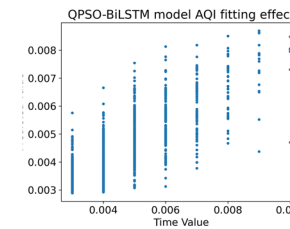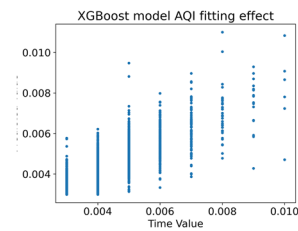(a) Comparison of different AQI models
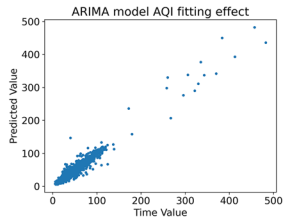
(b) ARIMA
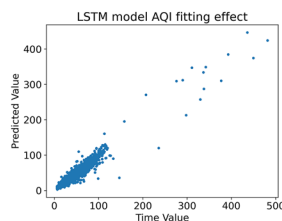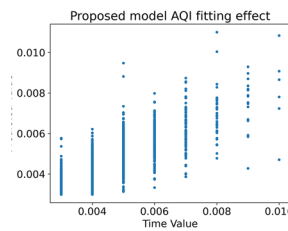

(c) LSTM
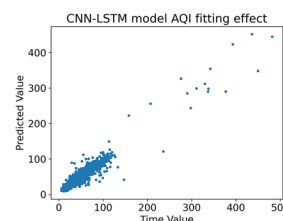
(d) CNN-LSTM
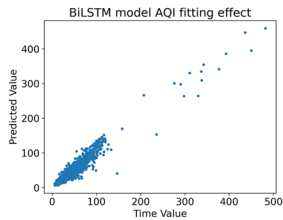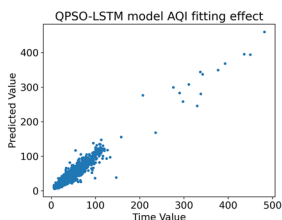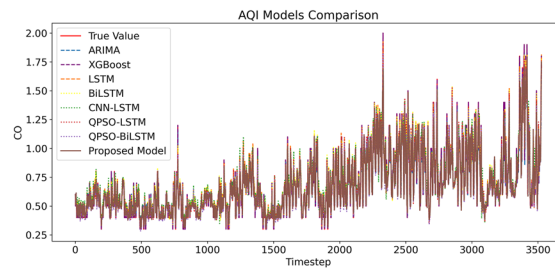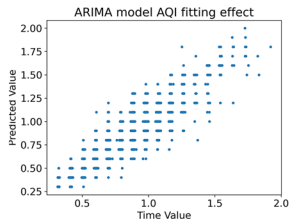

(e) BiLSTM

(f) QPSO-LSTM
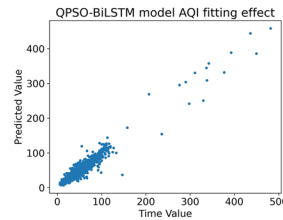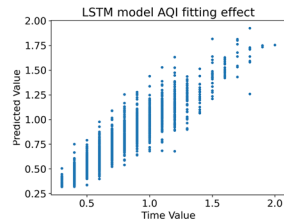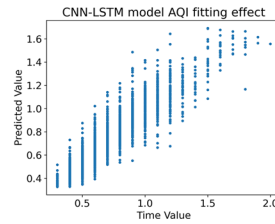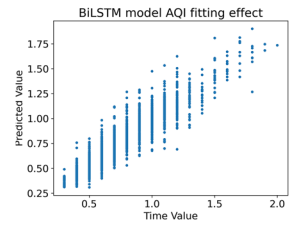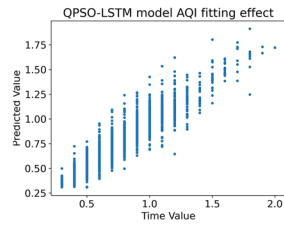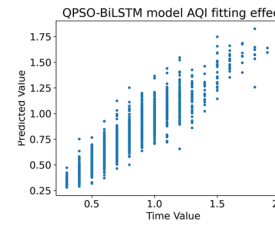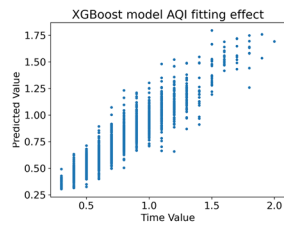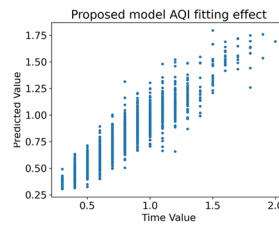

(g) QPSO-BiLSTM
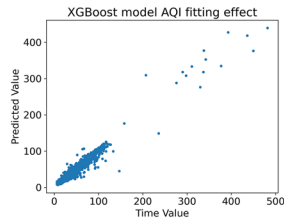
(h) XGBoost
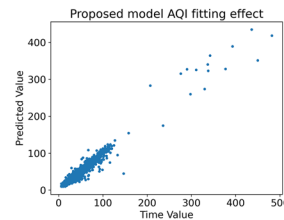
(i) Proposed Model

Fig. 18 PM_{10} concentration forecast comparison

(a) Comparison of different AQI models

(b) ARIMA

(c) LSTM
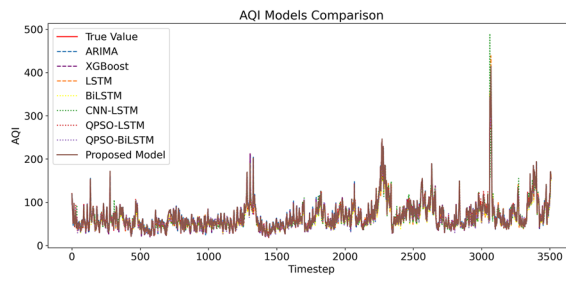

(d) CNN-LSTM
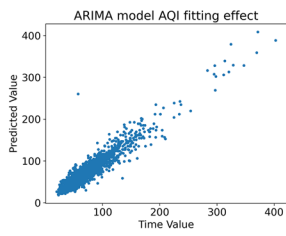

(e) BiLSTM

(f) QPSO-LSTM

(g) QPSO-BiLSTM

(h) XGBoost

(i) Proposed Model
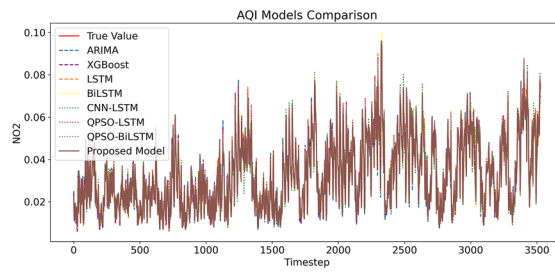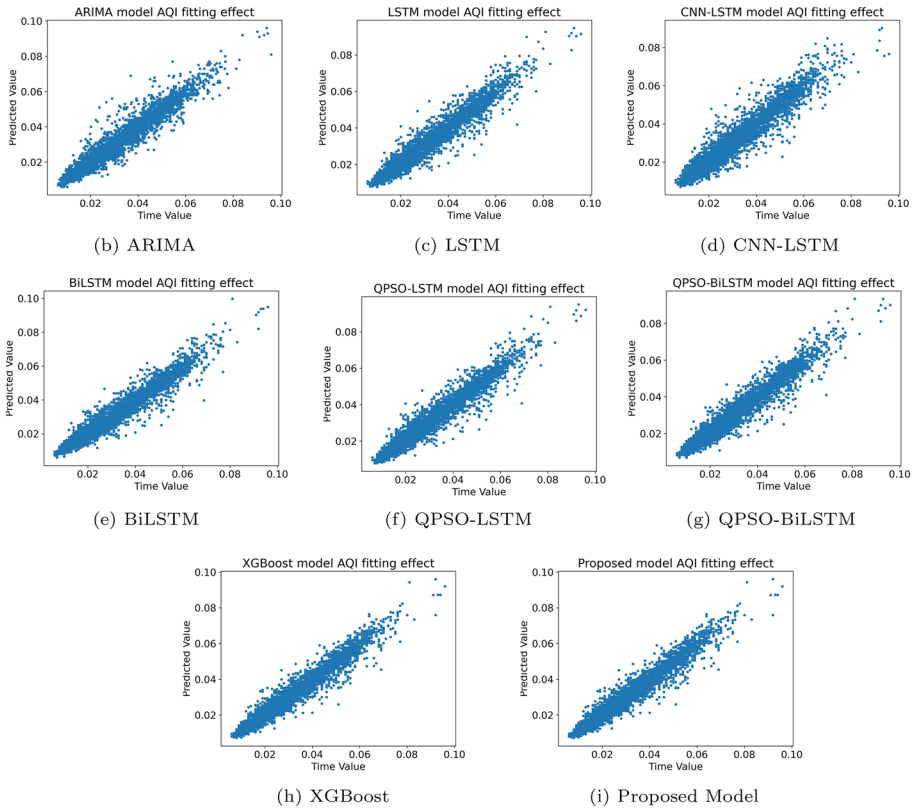
Fig. 19 NO_2 concentration forecast comparison

(a) Comparison of different AQI models

(b) ARIMA

(c) LSTM

(d) CNN-LSTM

(e) BiLSTM
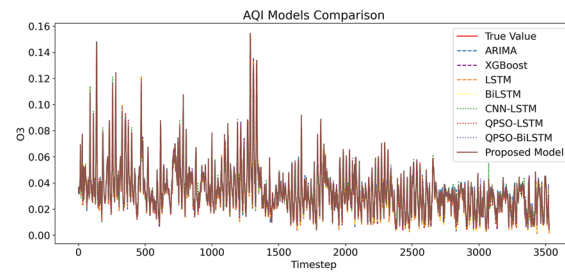
(f) QPSO-LSTM
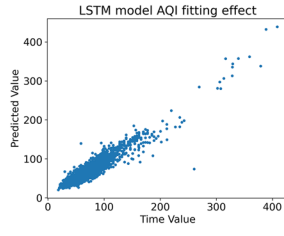

(g) QPSO-BiLSTM

(h) XGBoost
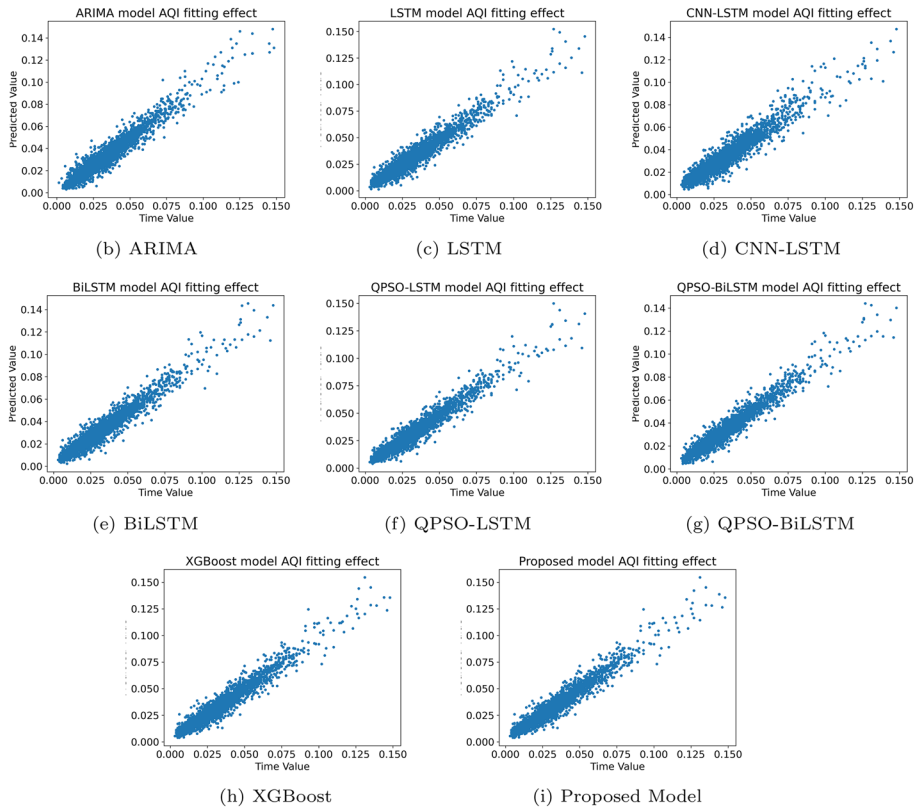
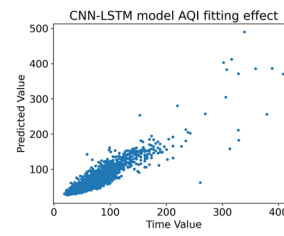(i) Proposed Model

Fig. 20 O_3 concentration forecast comparison
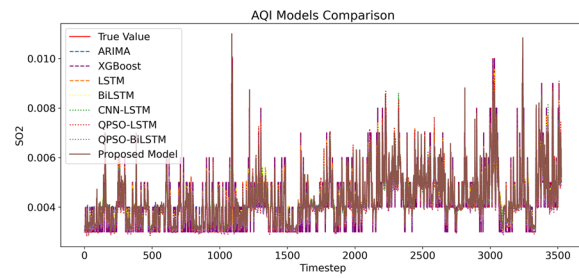
(a) Comparison of different AQI models
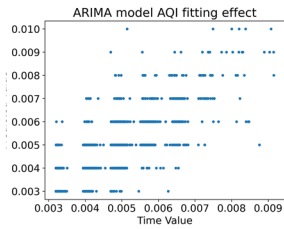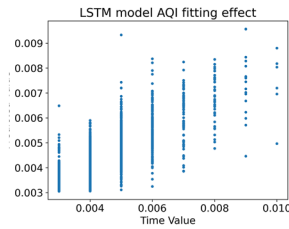
(b) ARIMA

(c) LSTM

(d) CNN-LSTM

(e) BiLSTM

(f) QPSO-LSTM

(g) QPSO-BiLSTM

(h) XGBoost

(i) Proposed Model

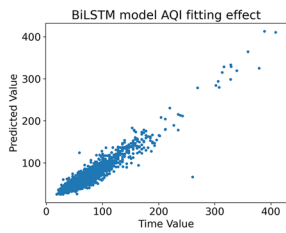**Fig. 21** SO_2 concentration forecast comparison

(a) Comparison of different AQI models

(b) ARIMA

(c) LSTM

(d) CNN-LSTM

(e) BiLSTM

(f) QPSO-LSTM

(g) QPSO-BiLSTM

(h) XGBoost

(i) Proposed Model

Fig. 22 CO concentration forecast comparison

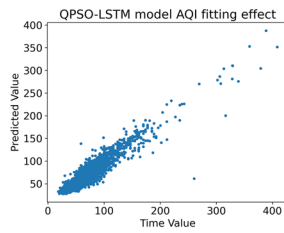(a) Comparison of different AQI models

(b) ARIMA

(c) LSTM

(d) CNN-LSTM

(e) BiLSTM

(f) QPSO-LSTM

(g) QPSO-BiLSTM

(h) XGBoost

(i) Proposed Model

Fig. 23 AQI forecasting results in Nowon-gu stations
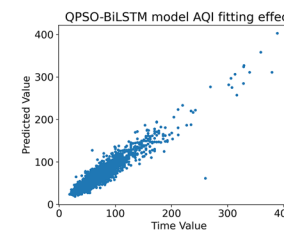
(a) Comparison of different AQI models
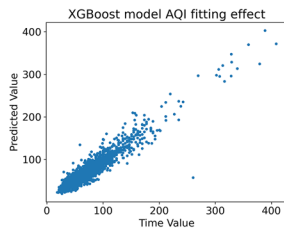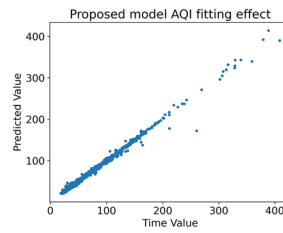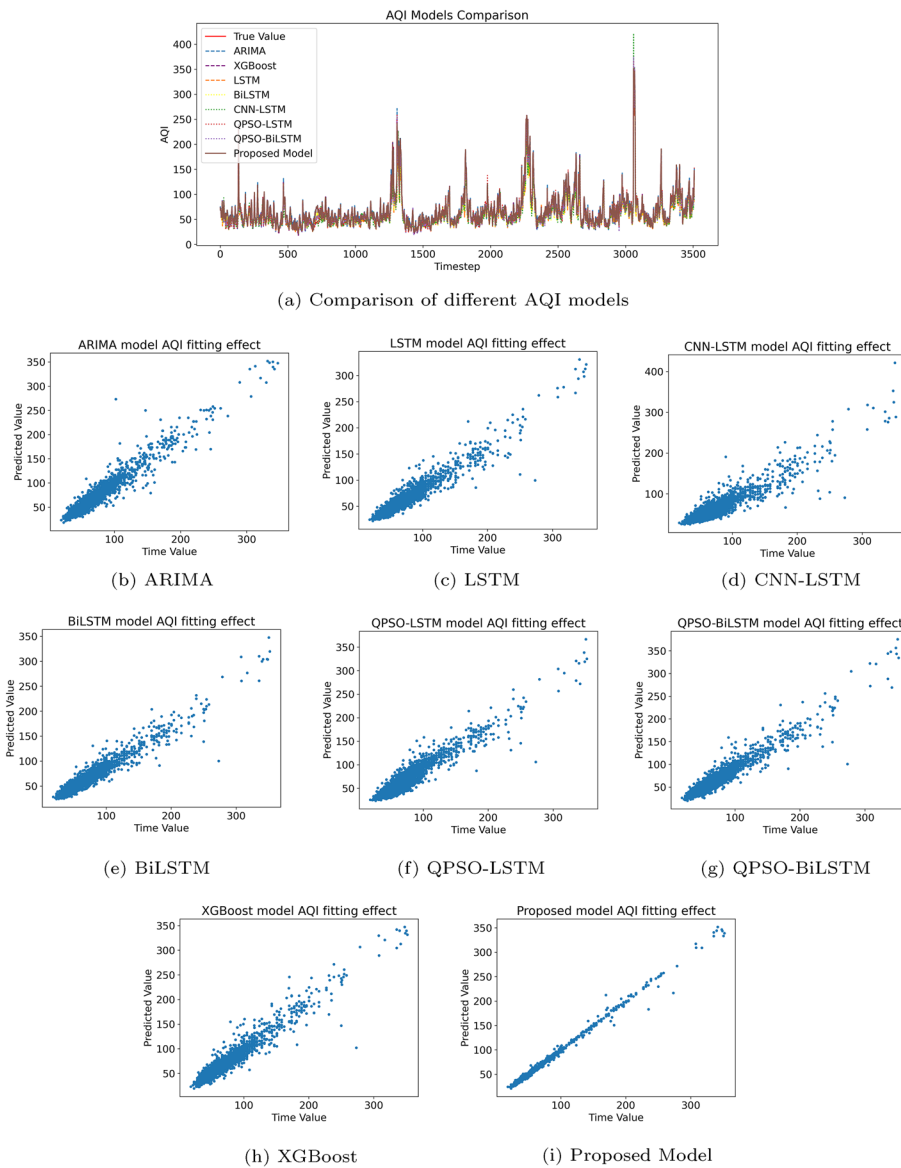
(b) ARIMA

(c) LSTM

(d) CNN-LSTM

(e) BiLSTM

(f) QPSO-LSTM

(g) QPSO-BiLSTM

(h) XGBoost

(i) Proposed Model

Fig. 24 AQI forecasting results in Dongjak-gu stations

Author contributions
All the authors contributed to the structuring of this paper, providing critical feedback and helping shape the research, analysis, and manuscript. AN conceived the presented idea, organized the manuscript and wrote the manuscript with input from all authors and implemented and tested the methodology. All the authors were involved in planning the work and supervised and reviewed the structure and contents of the paper.

Data availability
Sequence data that support the findings of this study have been deposited in the Seoul Air Pollution Data (open source) https://data.seoul.go.kr/.

Declarations

References

1. Zhu S, Lian X, Liu H, Hu J, Wang Y, Che J. Daily air quality index forecasting with hybrid models: a case in china. Environ Pollut. 2017;231:1232–44.
2. Lamichhane DK, Kim H-C, Choi C-M, Shin M-H, Shim YM, Leem J-H, Ryu J-S, Nam H-S, Park S-M. Lung cancer risk and residential exposure to air pollution: a Korean population-based case–control study. Yonsei Med J. 2017;58(6):1111.
3. Ahn H, Lee J, Hong A. Urban form and air pollution: clustering patterns of urban form factors related to particulate matter in Seoul, Korea. Sustain Cities Soc. 2022;81: 103859.
4. Zou B, You J, Lin Y, Duan X, Zhao X, Fang X, Campen MJ, Li S. Air pollution intervention and life-saving effect in china. Environ Int. 2019;125:529–41.
5. Jo H, Kim S-A, Kim H. Forecasting the reduction in urban air pollution by expansion of market shares of eco-friendly vehicles: a focus on Seoul, Korea. Int J Environ Res Public Health. 2022;19(22):15314. https://doi.org/10.3390/ijerph192215314.
6. Koo Y-S, Kim S-T, Cho J-S, Jang Y-K. Performance evaluation of the updated air quality forecasting system for Seoul predicting PM10. Atmos Environ. 2012;58:56–69.
7. AirKorea. https://airkorea.or.kr/. Accessed 31 Aug 2023.
8. Carbajal-Hernández JJ, Sánchez-Fernández LP, Carrasco-Ochoa JA, Martínez-Trinidad JF. Assessment and prediction of air quality using fuzzy logic and autoregressive models. Atmos Environ. 2012;60:37–50.
9. Zhang L, Tian X, Zhao Y, Liu L, Li Z, Tao L, Wang X, Guo X, Luo Y. Application of nonlinear land use regression models for ambient air pollutants and air quality index. Atmos Pollut Res. 2021;12(10): 101186.
10. Zhao L, Li Z, Qu L. Forecasting of Beijing PM2.5 with a hybrid ARIMA model based on integrated AIC and improved GS fixed-order methods and seasonal decomposition. Heliyon. 2022;8(12): e12239.
11. Zhou W, Wu X, Ding S, Cheng Y. Predictive analysis of the air quality indicators in the Yangtze river delta in China: an application of a novel seasonal grey model. Sci Total Environ. 2020;748: 141428.
12. Mehmood K, Bao Y, Cheng W, Khan MA, Siddique N, Abrar MM, Soban A, Fahad S, Naidu R, et al. Predicting the quality of air with machine learning approaches: current research priorities and future perspectives. J Clean Prod. 2022;379: 134656.
13. Mahalingam U, Elangovan K, Dobhal H, Valliappa C, Shrestha S, Kedam G. A machine learning model for air quality prediction for smart cities. In: 2019 international conference on wireless communications signal processing and networking (WiSPNET). 2019. p. 452–7. https://doi.org/10.1109/WiSPNET45539.2019.9032734.
14. Elsheikh AH. Applications of machine learning in friction stir welding: prediction of joint properties, real-time control and tool failure diagnosis. Eng Appl Artif Intell. 2023;121: 105961. https://doi.org/10.1016/j.engappai.2023.105961.
15. Ke H, Gong S, He J, Zhang L, Cui B, Wang Y, Mo J, Zhou Y, Zhang H. Development and application of an automated air quality forecasting system based on machine learning. Sci Total Environ. 2022;806: 151204.
16. Zhang W, Wu Y, Calautit JK. A review on occupancy prediction through machine learning for enhancing energy efficiency, air quality and thermal comfort in the built environment. Renew Sustain Energy Rev. 2022;167: 112704.
17. Gu Y, Li B, Meng Q. Hybrid interpretable predictive machine learning model for air pollution prediction. Neurocomputing. 2022;468:123–36.
18. Rakholia R, Le Q, Ho BQ, Vu K, Carbajo RS. Multi-output machine learning model for regional air pollution forecasting in ho chi Minh City, Vietnam. Environ Int. 2023;173: 107848.
19. Janarthanan R, Partheeban P, Somasundaram K, Elamparithi PN. A deep learning approach for prediction of air quality index in a metropolitan city. Sustain Cities Soc. 2021;67: 102720.
20. Zhang B, Rong Y, Yong R, Qin D, Li M, Zou G, Pan J. Deep learning for air pollutant concentration prediction: a review. Atmos Environ. 2022;290: 119347.
21. Saez M, Barceló MA. Spatial prediction of air pollution levels using a hierarchical Bayesian spatiotemporal model in Catalonia, Spain. Environ Model Softw. 2022;151: 105369.
22. Jurado X, Reiminger N, Benmoussa M, Vazquez J, Wemmert C. Deep learning methods evaluation to predict air quality based on computational fluid dynamics. Expert Syst Appl. 2022;203: 117294.
23. Zhou X, Xu J, Zeng P, Meng X. Air pollutant concentration prediction based on GRU method. J Phys Conf Ser. 2019;1168: 032058.
24. Mao W, Wang W, Jiao L, Zhao S, Liu A. Modeling air quality prediction using a deep learning approach: method optimization and evaluation. Sustain Cities Soc. 2021;65: 102567.
25. Elsheikh AH, Katekar VP, Muskens OL, Deshmukh SS, Elaziz MA, Dabour SM. Utilization of LSTM neural network for water production forecasting of a stepped solar still with a corrugated absorber plate. Process Saf Environ Prot. 2021;148:273–82. https://doi.org/10.1016/j.psep.2020.09.068.

26. Djouider F, Elaziz MA, Alhawsawi A, Banoqitah E, Moustafa EB, Elsheikh AH. Experimental investigation and machine learning modeling using LSTM and special relativity search of friction stir processed AA2024/Al$_2$O$_3$ nanocomposites. J Market Res. 2023;27:7442–56. https://doi.org/10.1016/j.jmrt.2023.11.155.

27. Wu Q, Lin H. A novel optimal-hybrid model for daily air quality index prediction considering air pollutant factors. Sci Total Environ. 2019;683:808–21.

28. Sarkar N, Gupta R, Keserwani PK, Govil MC. Air quality index prediction using an effective hybrid deep learning model. Environ Pollut. 2022;315: 120404.

29. Gilik A, Ogrenci AS, Ozmen A. Air quality prediction using CNN+ LSTM-based hybrid deep learning architecture. Environ Sci Pollut Res. 2022;29:1–19.

30. Rahman MM, Paul KC, Hossain MA, Ali GGMN, Rahman MS, Thill J-C. Machine learning on the covid-19 pandemic, human mobility and air quality: a review. IEEE Access. 2021;9:72420–50. https://doi.org/10.1109/ACCESS.2021.3079121.

31. Chang Y-S, Abimannan S, Chiao H-T, Lin C-Y, Huang Y-P. An ensemble learning based hybrid model and framework for air pollution forecasting. Environ Sci Pollut Res. 2020;27:38155–68.

32. Wang J, Li J, Wang X, Wang J, Huang M. Air quality prediction using CT-LSTM. Neural Comput Appl. 2021;33:4779–92.

33. Elsheikh AH, Saba AI, Elaziz MA, Lu S, Shanmugan S, Muthuramalingam T, Kumar R, Mosleh AO, Essa FA, Shehabeldeen TA. Deep learning-based forecasting model for covid-19 outbreak in Saudi Arabia. Process Saf Environ Prot. 2021;149:223–33. https://doi.org/10.1016/j.psep.2020.10.048.

34. Dai H, Huang G, Zeng H, Yu R. Haze risk assessment based on improved PCA-MEE and ISPO-LightGBM model. Systems. 2022;10(6):263.

35. Saba AI, Elsheikh AH. Forecasting the prevalence of covid-19 outbreak in Egypt using nonlinear autoregressive artificial neural networks. Process Saf Environ Prot. 2020;141:1–8. https://doi.org/10.1016/j.psep.2020.05.029.

36. Mirjalili S, Mirjalili SM, Hatamlou A. Multi-verse optimizer: a nature-inspired algorithm for global optimization. Neural Comput Appl. 2016;27:495–513.

37. Heydari A, Majidi Nezhad M, Astiaso Garcia D, Keynia F, De Santoli L. Air pollution forecasting application based on deep learning model and optimization algorithm. Clean Technol Environ Policy. 2022;24:1–15.

38. Heidari AA, Mirjalili S, Faris H, Aljarah I, Mafarja M, Chen H. Harris hawks optimization: algorithm and applications. Futur Gener Comput Syst. 2019;97:849–72.

39. Du P, Wang J, Hao Y, Niu T, Yang W. A novel hybrid model based on multi-objective Harris hawks optimization algorithm for daily PM2.5 and PM10 forecasting. Appl Soft Comput. 2020;96: 106620.

40. Marini F, Walczak B. Particle swarm optimization (PSO). A tutorial. Chemom Intell Lab Syst. 2015;149:153–65.

41. Huang Y, Xiang Y, Zhao R, Cheng Z. Air quality prediction using improved PSO-BP neural network. IEEE Access. 2020;8:99346–53.

42. Rajabioun R. Cuckoo optimization algorithm. Appl Soft Comput. 2011;11(8):5508–18.

43. Sun W, Sun J. Daily PM2.5 concentration prediction based on principal component analysis and LSSVM optimized by cuckoo search algorithm. J Environ Manag. 2017;188:144–52.

44. Trojovský P, Dehghani M. A new optimization algorithm based on mimicking the voting process for leader selection. PeerJ Comput Sci. 2022;8:976. https://doi.org/10.7717/peerj-cs.976.

45. Abd Elaziz M, Zayed ME, Abdelfattah H, Aseeri AO, Tag-eldin EM, Fujii M, Elsheikh AH. Machine learning-aided modeling for predicting freshwater production of a membrane desalination system: a long-short-term memory coupled with election-based optimizer. Alex Eng J. 2024;86:690–703. https://doi.org/10.1016/j.aej.2023.12.012.

46. Xue J, Shen B. Dung beetle optimizer: a new meta-heuristic algorithm for global optimization. J Supercomput. 2023;79(7):7305–36.

47. Duan J, Gong Y, Luo J, Zhao Z. Air-quality prediction based on the ARIMA-CNN-LSTM combination model optimized by dung beetle optimizer. Sci Rep. 2023. https://doi.org/10.1038/s41598-023-36620-4.

48. Cheung Y-W, Lai KS. Lag order and critical values of the augmented Dickey–Fuller test. J Bus Econ Stat. 1995;13(3):277–80.

49. Graves A. Long short-term memory. Berlin: Springer; 2012. p. 37–45. https://doi.org/10.1007/978-3-642-24797-2_4.

50. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput. 1997;9(8):1735–80.

51. Sutskever I, Vinyals O, Le QV. Sequence to sequence learning with neural networks. In: Advances in neural information processing systems; 2014.

52. Luo L, Yang Z, Yang P, Zhang Y, Wang L, Lin H, Wang J. An attention-based BiLSTM-CRF approach to document-level chemical named entity recognition. Bioinformatics. 2017;34(8):1381–8. https://doi.org/10.1093/bioinformatics/btx761.

53. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I. Attention is all you need. 2017. https://doi.org/10.48550/arxiv.1706.03762.

54. Shi Z, Hu Y, Mo G, Wu J. Attention-based CNN-LSTM and XGBoost hybrid model for stock prediction. 2023. arXiv: 2204.02623.

55. Wang D, Tan D, Liu L. Particle swarm optimization algorithm: an overview. Soft Comput. 2018;22:387–408. https://doi.org/10.1007/s00500-016-2474-6.

56. Sun J, Feng B, Xu W. Particle swarm optimization with particles having quantum behavior. In: Proceedings of the 2004 congress on evolutionary computation (IEEE Cat. No. 04TH8753), vol. 1. 2004. p. 325–3311. https://doi.org/10.1109/CEC.2004.1330875.

57. Mikki SM, Kishk AA. Quantum particle swarm optimization for electromagnetics. IEEE Trans Antennas Propag. 2006;54(10):2764–75. https://doi.org/10.1109/TAP.2006.882165.

58. Fang W, Sun J, Ding Y, Wu X, Xu W. A review of quantum-behaved particle swarm optimization. IETE Tech Rev. 2010;27(4):336–48. https://doi.org/10.4103/0256-4602.64601.

59. Zhao L, Cao N, Yang H. Forecasting regional short-term freight volume using QPSO-LSTM algorithm from the perspective of the importance of spatial information. Math Biosci Eng. 2023;20(2):2609–27.

60. Xu D, Zhang Q, Ding Y, Zhang D. Application of a hybrid ARIMA-LSTM model based on the SPEI for drought forecasting. Environ Sci Pollut Res. 2022;29(3):4128–44.

61. Abebe M, Noh Y, Kang Y-J, Seo C, Kim D, Seo J. Ship trajectory planning for collision avoidance using hybrid ARIMA-LSTM models. Ocean Eng. 2022;256: 111527.
62. Yin W, Schütze H, Xiang B, Zhou B. ABCNN: attention-based convolutional neural network for modeling sentence pairs. Trans Assoc Comput Linguist. 2016;4:259–72. https://doi.org/10.1162/tacl_a_00097.
63. Chen T, Guestrin C. Xgboost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. KDD '16. New York: Association for Computing Machinery; 2016. pp. 785–94. https://doi.org/10.1145/2939672.2939785.
64. Seoul air pollution data. https://data.seoul.go.kr/.
65. Altman EI, Iwanicz-Drozdowska M, Laitinen EK, Suvas A. Financial distress prediction in an international context: a review and empirical analysis of Altman's Z-score model. J Int Financial Manag Account. 2017;28(2):131–71. https://doi.org/10.1111/jifm.12053.
66. Das A, Ajila SA, Lung C-H. A comprehensive analysis of accuracies of machine learning algorithms for network intrusion detection. In: Machine learning for networking: second IFIP TC 6 international conference, MLN 2019, Paris, France, December 3–5, 2019, Revised Selected Papers 2. Springer; 2020. p. 40–57.
67. Kingma DP, Ba J. Adam: a method for stochastic optimization. 2017. arXiv:1412.6980.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.