

Received 12 July 2024, accepted 9 September 2024, date of publication 13 September 2024, date of current version 24 September 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3460187

RESEARCH ARTICLE

Multi-Site Air Quality Index Forecasting Based on Spatiotemporal Distribution and PatchTST-Enhanced: Evidence From Hebei Province in China

WENYI CAO^{ID1}, RUFERI ZHANG^{ID1,2}, AND WENXIN CAO^{ID3}

¹School of Economics, Hebei GEO University, Shijiazhuang 050031, China

²Hebei Center for Ecological and Environmental Geology Research, Hebei GEO University, Shijiazhuang 050031, China

³School of Foreign Languages, Southwest Jiaotong University, Chengdu 610031, China

Corresponding author: Rufei Zhang (zhangrufei@foxmail.com)

This work was supported in part by the Open Project Program of Hebei Center for Ecological and Environmental Geology Research under Grant JSYF-202305, and in part by the Science and Technology Program of Hebei Province under Grant 23557622D.

ABSTRACT Efficient and accurate air quality forecasting contributes significantly to environmental governance, health promotion, and the development of smart cities. However, few existing models can achieve multi-scale feature fusion and long sequence time series modeling in the prediction process. This study proposes a PatchTST-Enhanced model for multi-site air quality forecasting based on spatiotemporal distribution. It uses daily air quality data from 11 prefecture-level cities in Hebei Province from December 2, 2013, to October 12, 2023, to train the model. The new model demonstrates robust performance in Hebei's air quality forecasting (PreLen = 96: MSE = 0.5408, MAE = 0.5408, RSE = 0.5408; PreLen = 192: MSE = 0.2795, MAE = 0.2795, RSE = 0.2795; PreLen = 336: MSE = 0.6779, MAE = 0.6779, RSE = 0.6779; PreLen = 720: MSE = 0.6779, MAE = 0.6779, RSE = 0.6779), and the prediction accuracy has been significantly improved compared to both the pre-optimization model and other existing models. The PatchTST-Enhanced outperforms the PatchTST and improves it through four optimization modules: CGAttention, SiLU activation, AdamW optimizer, and SmoothL1 Loss function. By incorporating spatiotemporal features, the PatchTST-Enhanced can address the challenge of combining spatial and large temporal scales in air quality forecasting. The results provide critical information to protect health and improve the environment for the public.

INDEX TERMS Air quality forecast, atmospheric pollutants, spatiotemporal distribution, PatchTST-enhanced, deep learning.

I. INTRODUCTION

In the age featured by urbanization and industrialization, human demand for industrial materials and fossil fuels has increased dramatically. Air pollution in many cities has intensified. Previous research has shown that poor air quality has a detrimental influence on the health of residents. Exposure to air pollution environment can lead to approximately 6 units of lung capacity loss, and after long-term

The associate editor coordinating the review of this manuscript and approving it for publication was Abedalrhman Alkhateeb^{ID}.

exposure to such environment, the probability of suffering poor overall health status increases by 4.3% [1]. A significant positive correlation between atmospheric pollutants and heart failure-related hospitalizations has been revealed, and the risk of heart failure can be effectively reduced by controlling atmospheric pollutants [2]. Meanwhile, atmospheric pollutants also impose severe impacts on the immune and nervous systems, potentially leading to chronic diseases [3]. As of 2016, approximately 7 million global deaths were attributed to air pollution-related diseases, including cardiovascular diseases, cancer, stroke, and chronic obstructive pulmonary

disease (COPD) [4], [5]. Given the widespread concern and government initiatives to improve air quality, it is indispensable to propose precise and effective air quality forecasting models for air quality management and intervention.

Forecasting air quality is a crucial step for early intervention and control of urban air pollution [6]. Air Quality Index (AQI) provides a unified standard for assessing air quality. AQI synthesizes several atmospheric pollutant concentrations into a single conceptualized index value, showing the short-term air quality status and trend [7], [8]. The components of AQI are complex, and their changes are primarily influenced by pollution emissions, including particulate matter 2.5 (PM2.5), particulate matter 10 (PM10), sulfur dioxide (SO₂), carbon monoxide (CO), nitrogen dioxide (NO₂), and ozone (O₃) [9]. Besides, meteorological conditions, terrain, regional position, and social and economic are also significant factors affecting air quality [10]. Under natural conditions, changes in meteorological factors such as temperature, air pressure, wind speed and direction can directly alter the concentration of atmospheric pollutants. Factors like light, visibility, and humidity can also produce photocatalytic reactions with atmospheric pollutants [11], [12], [13]. These factors all influence the concentration and spatiotemporal distribution characteristics of atmospheric pollutants to a certain extent. Therefore, it is difficult to directly observe the spatiotemporal distribution characteristics of air quality in various regions and achieve accurate forecasting using existing data from regional environmental monitoring stations [14]. However, historical data on spatiotemporal relationships, meteorological characteristics, and atmospheric pollutants can provide a basis for model training and learning to predict future air quality [15], [16].

Hebei Province, a significant economic region in P. R. China, is located between 36°05'N and 42°40'N latitude, 113°27'E and 119°50'E longitude. Due to its geographical location, the region suffers from severe air pollution caused by its dry climate and minimal wind, which hinder the dispersion of pollutants [17], [18]. In recent years, air pollution has been further aggravated by urbanization and the increasing car ownership [19]. Furthermore, given the province's abundant ecological resources (e.g. natural scenic spots, forests, and wetlands), improving air quality is crucial for maintaining ecological balance, protecting biodiversity, and stabilizing ecosystems [20]. Additionally, Hebei Province is one of the most densely populated provinces in China. Combating air pollution in this region is essential for residents' health [21]. Finally, as one of the economically developed regions in China, the environmental quality of Hebei Province directly impacts regional cooperation and international image. Considering its typicality and importance as a region facing severe air pollution issues, Hebei Province serves as an ideal candidate for air quality forecasting research, but there have been few studies on air quality forecasting in Hebei Province. Therefore, the present research bears several practical implications.

Generally, the methods for air quality forecast can be grouped into three categories: statistical model (ARIMA, regression analysis, time series analysis), physical model (WRF-weather research and forecasting, CMAQ-community multiscale air quality) and machine learning algorithm (deep learning, random forest, SVM) [22], [23], [24], [25], [26], [27], [28]. These diverse methods are all effective in forecasting air quality, sharing a common focus on addressing the spatiotemporal correlation in the prediction process. Therefore, during the algorithm development, the correlation between spatial proximity and temporal sequence should be taken into account to enhance the accuracy of air quality forecasting. Deep learning prediction algorithms possess robust data processing and learning capabilities, which not only enable them to achieve precise predictions through effective processing of large-scale data, but also facilitate the learning of abstract features from spatiotemporal data, thus better distinguishing and predicting diverse spatiotemporal patterns [29], [30], [31]. Air quality forecasting involves complex relationships among multiple variables and factors, including historical observation data, meteorological data, and geospatial data [32], [33], [34]. Therefore, deep learning prediction algorithms are one of the best methods for achieving accurate air quality forecasting.

The Transformer algorithm employs a self-attention mechanism, which is ideal for spatiotemporal forecasting as it captures long-term dependencies across different temporal axes and spatial ranges in processing time series data [35], [36], [37]. However, most scholars' optimization strategies for Transformers predominantly focus on the modulation of attention mechanisms, without delving into the integration of various optimizers or loss functions. Moreover, two major issues arise when utilizing the Transformer directly for time series forecasting: Firstly, the masking values at the current time step can be easily inferred by interpolating with the immediate proceeding or succeeding time values without high level understanding of the entire sequence, contradicting our objective of learning critical abstract representations of the entire signal. Secondly, given L time steps, D-dimensional space, M variables, and each variable having a prediction horizon T, the output requires a parameter matrix with dimensions of (L·D) × (M·T). If either one or all of these four values are large, this matrix can become exceptionally oversized, potentially leading to overfitting when downstream training samples are scarce [38]. Subsequently, Nie et al. [38] proposed a novel modal based on Transformer for time series forecasting and representation learning, converting time series data into forms similar to Patch in Vision Transformer, called PatchTST (patch time series Transformer). This modal masks data blocks rather than individual data points, and forecasts entire patches, thereby overcoming the aforementioned issues and significantly enhancing model performance [38], [39], [40], [41].

Inspired by the new modal, this research aims to improve the algorithm based on PatchTST and realize spatiotemporal

feature fusion. Since PatchTST is a newly developed algorithm in 2023, few scholars have optimized it yet. In this paper, we propose a PatchTST- Enhanced algorithm that enhances PatchTST through Cascaded Group Attention (CGAttention), Sigmoid Linear Unit (SiLU), Adam Weight Decay Optimizer (AdamW), and SmoothL1 Loss. CGAttention can flexibly adjust the number and parameter settings of attention layers according to the specific task requirements and data characteristics to achieve optimal performance [42]. The input variables for air quality forecasting involve multiple features such as time, space, meteorology, and atmospheric pollutants. PatchTST-CGAttention can extract higher-level feature representations step by step from the input data, thereby improving the forecasting accuracy of the model [43], [44]. This step-by-step feature extraction helps the model better understand the intrinsic structure and patterns of complex air quality forecasting datasets [45]. Subsequently, we attempt to replace the activation function of the PatchTST-CGAttention feature extraction network with SiLU, which helps avoid gradient disappearance or gradient explosion during training, thus accelerating the convergence process and improving the training stability of the model [46]. On this basis, AdamW algorithm is introduced and the original loss function is replaced with SmoothL1 Loss, which can help to prevent overfitting, improve the generalization ability of the model, avoid the discontinuity of gradients in multi-scale feature data, and ensure the model robustness when outliers exist [47], [48], [49], [50]. Finally, comparative and ablation experiments are conducted on the air quality dataset of Hebei Province based on PatchTST-Enhanced. The results of comparative experiments demonstrate that the proposed PatchTST-Enhanced outperforms other models, while ablation experiments prove the necessity and effectiveness of enhancing each module of PatchTST [51], [52].

The objective of this research is to establish a multi-site deep learning algorithm (PatchTST-Enhanced) based on spatiotemporal distribution to forecast the AQI in Hebei province, P. R. China. The innovation of this paper is three-fold. Firstly, a new PatchTST-Enhanced algorithm is proposed, which enhances PatchTST algorithm with CGAttention, SiLU, AdamW, SmoothL1 Loss. When trained on the dataset in this research, PatchTST-Enhanced algorithm demonstrates improvements in both accuracy and model performance compared to the traditional PatchTST algorithm. Secondly, ablation experiments are conducted through four optimization schemes. After evaluating the performance of various combinations of optimized PatchTST-Enhanced algorithms, the optimal algorithm suitable for air quality forecasting in Hebei Province is selected, which achieves excellent performance on four different lag orders. Lastly, by integrating spatiotemporal features into the PatchTST- Enhanced algorithm, the challenge of combining regional and large temporal scales in air quality forecasting can be addressed. This forecasting model aligns better with the spatiotemporal variations of atmospheric pollutants, enhancing its applicability and accuracy.

Overall, the most significant contribution of this research lies in proposing a deep learning-based air quality forecasting algorithm tailored for Hebei Province, P. R. China. This algorithm is vital in providing timely information and early warnings, thereby assisting individuals and relevant authorities in making appropriate decisions and taking actions, so as to ensure residents' health, improve the environment, and foster sustainable development.

This paper is organized as follows: Section I presents the introduction. Section II reviews the related work on air quality forecasting and deep learning prediction. Section III describes the deploy of PatchTST model and the model optimization scheme to improve the accuracy of the air quality forecasting in Hebei Province. Section IV describes the study area of the research and the data used in this paper. Section V applies the above method to train the datasets and discuss the results. Finally, Section VI is the conclusion.

II. RELATED WORK

Zhang et al. [53] examined the factors that might affect people's health through geographically weighted regression (GWR) and found that air quality is one of the key factors. In recent years, the United Nations Environment Programme (UNEP) and the World Health Organization (WHO) have been working to promote air quality improvement on a global scale. Many countries and regions have adopted more stringent environmental protection policies and regulations to reduce air pollution and improve air quality. In the new context, research on air quality has gained the attention of a wide range of scholars [54], [55], [56]. Zhang et al. [14] identified air pollutants in 74 cities in China from the perspective of spatial and seasonal distribution and found that the distribution of atmospheric pollutants has obvious spatial correlation and seasonal effects. Suris et al. [57] used a machine learning approach to cluster the PM10 in Malaysia and found that the clustering of the atmospheric pollutants was heavily influenced by its location and geographical area. Zhou et al. [58] used the non-radial DEA model and the PS (Phillips and Sul) convergence test to study the spatial distribution of atmospheric pollutants and found that there was a significant unbalanced distribution of atmospheric pollutants from different regions. Based on previous research, the study of air quality may need to take into account spatial distribution and the fusion of spatial and temporal features.

Air quality forecasting has always been important in air quality research. The results can be used to assess the effectiveness of different environmental interventions, analyze pollution sources, and evaluate the effectiveness of policy implementation, etc [59], [60], [61]. Previous research on air quality forecasting always focused on traditional time series methods or machine learning. Kumar and Jain [62] predicted the daily average air pollutants at various transportation stations in New Delhi city through the stationary stochastic ARMA/ARIMA model, and the results of the model predictions were judged to be satisfactory as compared to the previous studies. Sun and Sun [63] designed

a model improvement scheme for the Least Squares Support Vector Machine (LSSVM) according to Cuckoo Search (CS) and used it for PM2.5 concentration prediction studies. This model demonstrated superior performance in air quality forecasting and largely improved the generalization ability of air quality forecasting models. Liu et al. [64] utilized an improved CMAQ-ARIMA model to predict PM2.5, NO₂, and O₃ at various stations in Hong Kong, and the RMSD of this model could be reduced by 14.3-21.0%, 41.2-46.3%, and 47.8-49.7%, respectively, compared with the traditional CMAQ model. Xiong et al. [65] proposed an improvement of the gray prediction model based on iterative reweighted least squares (IRLS) and used it for the air quality forecasting in Shanghai, China, which achieved an improvement of the model performance compared with the traditional gray prediction model. However, the traditional prediction models still had some shortcomings in the field of air quality prediction. Zhu et al. [66] pointed out that air pollutant data were often non-stationary and chaotic, so it was very difficult to conduct accurate predictions. Moreover, most multivariate linear models, autoregressive integrated moving average (ARIMA) and support vector regression (SVR), etc., were unable to fully learn all the characteristics of complex pollution indicators.

In recent research, air quality forecasting with spatiotemporal feature fusion is a hot issue. The prediction process needs to involve multi-scale features, such as sensor data, meteorological data, demographic data, etc. Integrating the features of spatial and temporal dimensions, deep learning algorithms are one of the most effective methods to realize the prediction of multi-scale feature data [67], [68], [69]. Yan et al. [10] predicted the air quality index at multiple stations in Beijing based on spatiotemporal distribution and using BPNN, convolutional neural network (CNN), long short-term memory (LSTM) model and CNN-LSTM multiple deep learning algorithms, and concluded that the CNN-LSTM model had a powerful performance in air quality forecasting. However, traditional neural networks may encounter the problem of vanishing gradient or exploding gradient when dealing with long-term dependencies [70]. In contrast, the self-attention mechanism in Transformer allows the model to compute the attention at each position in the sequence, thus better capturing long-term dependencies [71]. Furthermore, the Transformer can adaptively learn temporal dependencies and spatial correlations to better capture spatio-temporal features in sequences [72]. Alerskans et al. [73] implemented the automatic extraction of important input variables based on self-attention in the Transformer and predicted temperature successfully. The results showed that Transformer outperforms NWP (numerical weather prediction) in temperature prediction and other traditional models.

Initially, Transformer was mainly used in the field of computer vision and image segmentation. A wide range of scholars have continuously adapted the attention mechanism of Transformer to improve its performance in different scientific fields [74], [75]. In previous studies, Transformer

has been used in time series forecasting with certain shortcomings, and the forecasting accuracy often failed to exceed that of classical linear forecasting models such as DLlinear [76]. Nie et al. [38] introduced an approach to combine Patch modeling with Transformer, where the input time series were segmented into “patches” using fixed-size windows and strides, akin to image segmentation practices. These segmented patches were then utilized as inputs for the Transformer model, facilitating prediction through advanced modeling. The proposed PatchTST model, in contrast to previous works, can effectively capture local semantic information while also learning from extended backtracking windows, leading to enhanced prediction accuracy and model performance. Given the recent introduction of PatchTST, few scholars have conducted research on optimizing its modules thus far. Liu et al. [42] observed high redundancy among feature heads, resulting in computational inefficiency. To address this, they devised a cascaded group attention module, which can not only minimize computational costs but also bolster attention diversity, subsequently accelerating feature learning, model execution, and prediction precision. Inspired by these findings, we propose to substitute the original attention mechanism in PatchTST with the cascaded group attention module, aiming to further amplify the model’s capability in capturing long-distance dependencies - a pivotal aspect for temporal prediction, where intricate long-term relationships within time series data are prevalent.

In addition to adjustments in the attention mechanism, optimization modules (such as activation functions, optimizers, and loss functions) offer promising avenues for enhancing model performance [77], [78], [79]. Nishiyama et al. [80] found that the SiLU function, being smooth and continuously differentiable, incurred lower computational costs compared to complex activation functions, thereby mitigating the computational burden of Transformer models and accelerating both training and inference. Zhou et al. [48] demonstrated that by integrating weight decay, AdamW could effectively mitigate overfitting and enhance model generalization. Adam can optimize each parameter’s learning rate by estimating the first moment (mean) and second moment (uncentered variance) of gradients, often converging faster, particularly with sparse gradients [47]. Weight decay, a regularization technique, can prevent overfitting by appending a term proportional to the weight values to the loss function [81]. In contrast to traditional Adam, weight decay was applied separately. AdamW can seamlessly integrate it, facilitating a more concise and unified implementation [82]. Furthermore, the loss function is integral to the model optimization process [83]. SmoothL1 loss, a variant of L1 loss, employs L2 loss near zero to mitigate gradient discontinuity. It favors L2 loss for small errors, expediting convergence, and L1 loss for large errors, ensuring robustness against outliers. The robustness stems from the L1 loss, which handles outliers efficiently. SmoothL1 loss utilizes L2 loss for small errors, which accelerates convergence during optimization. Its implementation is straightforward and typically necessitates minimal

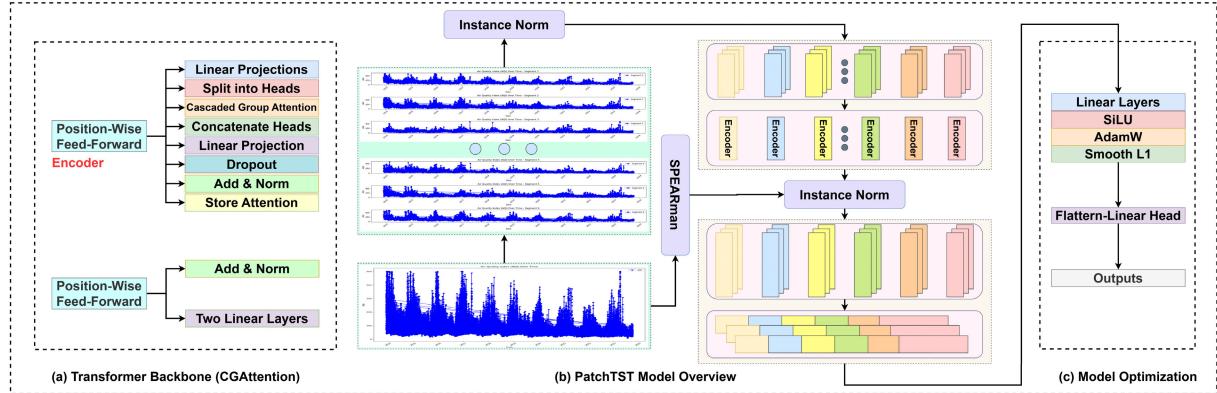


FIGURE 1. Architecture of PatchTST-Enhanced. (a) The backbone of Transformer encoder and the attention mechanism is replaced with cascaded group attention; (b) Multivariate time series data is divided into distinct channels, which share the same Transformer backbone, but the forward processes are independent. Each channel univariate series undergoes instance normalization operator, and is subsequently segmented into patches. These patches serve as Transformer input tokens; (c) SiLU, AdamW, and SmoothL1 Loss are used to enhance the PatchTST algorithm and output the results.

hyperparameter tuning [84]. Leveraging these optimization modules, we refine the previous PatchTST model, ultimately training the PatchTST-Enhanced model in this study.

III. METHODOLOGY

The primary objective of this research is to propose a PatchTST-Enhanced model for forecasting air quality in Hebei Province. To conduct comparative experiments, we also trained and employed GA-BP, CNN-LSTM, and PSO-SVM algorithms with the same datasets, thereby demonstrating the superiority of the proposed PatchTST-Enhanced algorithm over the current mainstream algorithms for air quality forecasting. The architecture of PatchTST-Enhanced is shown in Figure 1. The model uses Transformer encoder enhanced by CGAttention as its core architecture.

A. MODEL STRUCTURE

Firstly, the i -th univariate series of length L starting at time index 1 is denoted as Equation (1):

$$x_{1:L}^{(i)} = (x_1^{(i)}, \dots, x_L^{(i)}) \quad (1)$$

where $i = 1, \dots, N$. Then, the input (x_1, \dots, x_L) is split to N univariate series $x^{(i)} \in \mathbb{R}^{1 \times L}$, where each univariate series is fed independently into the Transformer backbone according to channel-independence setting. Transformer backbone will provide prediction results accordingly, as shown in Equation (2):

$$\hat{x}^{(i)} = (\hat{x}_{L+1}^{(i)}, \dots, \hat{x}_{L+T}^{(i)}) \in \mathbb{R}^{1 \times T} \quad (2)$$

According to the research of Nie et al. [38], each input univariate time series $x^{(i)}$ is divided into patches, which can be either overlapped or non-overlapped. The length of the patch is denoted as P , and the stride - the non-overlapping region between two consecutive patches is denoted as S . In the process of patching, a series of patches $x_p^{(i)} \in \mathbb{R}^{P \times N}$ will be generated, where N is the number of patches. The

Equation (3) is as follows:

$$N = \left\lfloor \frac{(L - P)}{S} \right\rfloor + 2 \quad (3)$$

Here, we pad S repeated numbers of the last value $x_L^{(i)} \in \mathbb{R}$ to the end of the original sequence before patching. After patching the original series, we use a vanilla Transformer encoder that maps the observed signals to the latent representations [85], [86]. The patches are mapped to the Transformer latent space of dimension D via a trainable linear projection $W_P \in \mathbb{R}^{D \times P}$, and a learnable additive position encoding $W_{pos} \in \mathbb{R}^{D \times N}$ is used to monitor the temporal order of patches as Equation (4):

$$x_d^{(i)} = W_P x_p^{(i)} + W_{pos} \quad (4)$$

where, $x_d^{(i)} \in \mathbb{R}^{D \times N}$ denote the input that will be fed into Transformer encoder in Figure 1 (a) and (b). Then, each head in $h = 1, \dots, H$ multi-head attention will transform them into query matrices $Q_h^{(i)} = (x_d^{(i)})^T W_h^Q$, key matrices $K_h^{(i)} = (x_d^{(i)})^T W_h^K$ and value matrices $V_h^{(i)} = (x_d^{(i)})^T W_h^V$, where $W_h^Q, W_h^K \in \mathbb{R}^{D \times d_k}$ and $W_h^V \in \mathbb{R}^{D \times D}$. Subsequently, a scaled production is employed to generate attention output $O_h^{(i)} \in \mathbb{R}^{D \times N}$ in Equation (5):

$$\begin{aligned} (O_h^{(i)})^T &= \text{Attention}(Q_h^{(i)}, K_h^{(i)}, V_h^{(i)}) \\ &= \text{Softmax}\left(\frac{Q_h^{(i)} K_h^{(i)T}}{\sqrt{d_k}}\right) \end{aligned} \quad (5)$$

B. MODEL OPTIMIZATION

Compared with the research of Nie et al. [38], we use an innovative cascaded group attention (CGA) module for PatchTST. The air quality data we collected contained information at different scales. CGA can divide the input features into distinct segments, each of which is fed into an attention head. Each head computes its self-attention mapping, then cascades

the outputs of all the heads and projects them back into the dimension of the input via a linear layer [42]. This attention mechanism is shown in Equation (6):

$$\begin{aligned}\tilde{X}_{ij} &= \text{Attn} \left(X_{ij} W_{ij}^Q, X_{ij} W_{ij}^K, X_{ij} W_{ij}^V \right), \\ \tilde{X}_{i+1} &= \text{Concat} \left[\tilde{X}_{ij} \right]_j = W_i^P \left[\cdot \right]_{1:h}\end{aligned}\quad (6)$$

where j -th head is used to compute the self-attention of X_{ij} . X_{ij} is the j -th split of the input feature X_i . $X_i = [X_{i1}, X_{i2}, \dots, X_{ih}]$ and $1 \leq j \leq h$. Denote the total number of heads as h . W_{ij}^Q , W_{ij}^K and W_{ij}^V are projection layers mapping the input feature split into different subspaces. W_i^P is a linear layer that projects the concatenated output features back to the dimension consistent with the input [42].

After replacing the attention mechanism in PatchTST with cascaded group attention, we can compute the attention map of each head according to cascaded manner. The output results from each head are stacked into subsequent heads, progressively refining the representation of each feature, as shown in Equation (7):

$$X'_{ij} = X_{ij} + \tilde{X}_{i(j-1)}, \quad 1 < j \leq h \quad (7)$$

X'_{ij} is the addition of the j -th input split X_{ij} and the $j - 1$ -th head output $\tilde{X}_{i(j-1)}$ calculated by Eq. 6. X'_{ij} will replace X_{ij} to serve as the new input feature for the j -th head when calculating the self-attention [42].

Then, the cascaded group attention is incorporated into PatchTST. To prevent PatchTST-CGAttention from failing to provide consistent relational predictions for positive and negative input values during complex feature extraction, a Sigmoid Gated Linear Unit (SiLU) function is incorporated into the model. SiLU exhibits notable smoothness and stronger derivability around zero, facilitating gradient calculation and update, thereby enhancing the performance of the feature extraction network. The expression of SiLU is presented in Equation (8):

$$\text{SiLU}(x) = x \times \left(\frac{1}{1 + e^{-x}} \right) \quad (8)$$

To mitigate the potential overfitting issues encountered during feature recognition, the integration of the SiLU function is followed by the adoption of the AdamW optimizer for training. This strategy effectively regulates model complexity, mitigates the risk of overfitting, and concurrently enhances the model's robustness [87]. The process of the AdamW optimizer is illustrated in Equations (9), (10), (11), (12):

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t \quad (9)$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \quad (10)$$

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t}, \quad \hat{v}_t = \frac{v_t}{1 - \beta_2^t} \quad (11)$$

$$\theta_{t+1} = \theta_t - \eta \left(\frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon} + w \theta_t \right) \quad (12)$$

where β_1 and β_2 are exponential decay time constants. m_t and v_t are first and second order momentum respectively. w is the decay factor. η is learning rate. g_t is the gradient of the parameters. θ_t is the model parameter that need to be learned through optimization.

During the training process utilizing the AdamW optimizer, the SmoothL1 Loss function is introduced. This loss function can mitigate the loss of critical information during feature learning and facilitate faster convergence of the PatchTST model during optimization. The SmoothL1 Loss function is formulated as shown in Equation (13):

$$L_{\text{SmoothL1}}(x, \beta) = \begin{cases} \frac{|x|^2}{2\beta}, & |x| < \beta \\ |x| - \frac{\beta}{2}, & \text{other} \end{cases} \quad (13)$$

where x is regression label, β is a hyper-parameter, and $L_{\text{SmoothL1}}(\cdot)$ serves as a crucial component in the localization loss function L_{loc} . The computation of L_{loc} is shown in Equation (14):

$$L_{\text{loc}} = \sum_{i \in \{x, y, w, h\}} w L_{\text{SmoothL1}}(t_i - v_i, \beta) \quad (14)$$

where L_{loc} is defined on the ground-truth bounding box of objective tuple. i denotes the index of bounding box. t_i is the i -th bounding box predictions. v_i is the i -th ground-truth bounding box. x and y in $\{x, y, w, h\}$ are the x -coordinate and y -coordinate of the bounding box center respectively. w and h are the width and height of the bounding box center, respectively.

After optimizing and enhancing the model, air quality forecasting for Hebei Province is conducted, accompanied by comparative and ablation experiments. The comparative experiments primarily demonstrate the model's performance by training on the same dataset using GA-BP, CNN-LSTM, and PSO-SVM models. The ablation experiments provide contrasting evidence through comparisons between the original PatchTST model and PatchTST-Enhanced models.

IV. STUDY AREA AND DATASET

A. STUDY AREA

This research selects Hebei Province, P. R. China, as the study area due to its status as one of the most severely air polluted regions in the country. As a significant economic hub and a densely populated province, Hebei Province is located in Northern China. Bordering the Bohai Sea to the east, Hebei exhibits a gradually flattened terrain from north to south, with undulating mountains in the west. Its unique geographical position underscores its significance in transportation, economics, and culture. Characterized by a typical temperate continental climate, Hebei experiences climatic variations across its vast territory. Precipitation is predominantly concentrated in summer, with relatively higher amounts in the southern regions. Influenced by the geographical location, climatic factors, intensive transportation networks, and the

concentration of industrial zones, atmospheric pollutants are not easily dispersed.

AQI represents the most direct method for quantifying regional air quality and is the most widely employed index in air quality forecasting [10]. Ma et al. [88] conducted a study based on relevant environmental quality assessment regulations and criteria, which pointed out that AQI calculations currently encompass six primary pollutant indicators: PM2.5, PM10, SO₂, CO, NO₂, and O₃. Currently, Hebei Province has several ambient air quality monitoring stations. For rapid and accurate assessment and forecasting of air quality, this research selects data from 11 prefecture-level city stations within the province, including Shijiazhuang, Xingtai, Handan, Zhangjiakou, Chengde, Qinhuangdao, Langfang, Tangshan, Baoding, Cangzhou, Hengshui. The study areas and distributions of the ambient air quality monitoring stations across these 11 prefecture-level cities in Hebei Province are depicted in Figure 2 and Table 1.

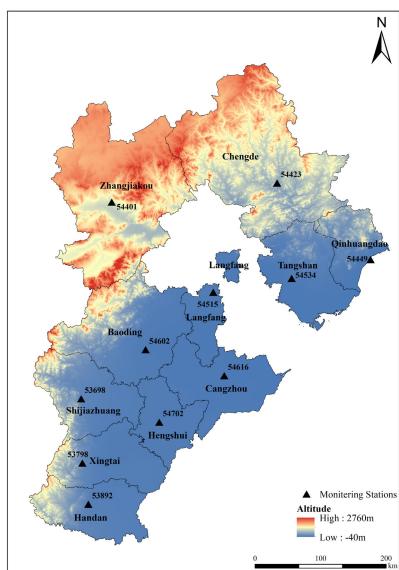


FIGURE 2. study area.

TABLE 1. Location information of air quality monitoring stations in prefecture-level cities of Hebei province.

Monitoring station	Longitude	Latitude	Prefecture-Level City
No. 53698	38.0666666	114.35	Shijiazhuang
No. 53798	37.183333	114.366667	Xingtai
No. 53892	36.616667	114.466667	Handan
No. 54401	40.766667	114.916667	Zhangjiakou
No. 54423	40.966667	117.916667	Chengde
No. 54449	39.85	119.516667	Qinhuangdao
No. 54515	39.5	116.7	Langfang
No. 54534	39.65	118.1	Tangshan
No. 54602	38.733333	115.483333	Baoding
No. 54616	38.35	116.85	Cangzhou
No. 54702	37.733333	115.7	Hengshui

B. DATASETS

This research constructs the datasets for model training across four dimensions: air quality, atmospheric pollutants data, meteorological data, and spatiotemporal data. The

air quality and atmospheric pollutants data are sourced from the China National Environmental Monitoring Centre (<http://www.cnemc.cn/sssj/>), encompassing daily AQI, PM2.5, PM10, SO₂, NO₂, O₃, and CO measurements from ambient air quality monitoring stations in 11 prefecture-level cities in Hebei Province, spanning from December 2, 2013, to October 12, 2023. The meteorological data is constructed by aggregating information from the China Meteorological Data Service Centre (<http://data.cma.cn/site/index.html>), encompassing daily records of temperature, pressure, and wind direction (quantified using numbers 1-17, adhering to the standardization outlined in Table 2), wind speed, visibility, rain, and snow for the 11 cities. Here, rain and snow are represented as two-valued variables, with 0 indicating absence and 1 indicating presence. The spatiotemporal data incorporates seasonal and quarterly variations, along with the latitude and longitude of each monitoring station, week number within the year, day of the week, and holiday status. Holiday designation is based on nationally recognized statutory holidays and weekends, with 1 representing holidays and 0 representing weekdays. In the process of training the forecasting model, AQI serves as the output variable, while all other factors mentioned constitute the input variables.

TABLE 2. Wind direction quantitative standard.

No.	Meaning	No.	Meaning	No.	Meaning
1	N	2	NNE	3	NE
4	ENE	5	E	6	ENE
7	SE	8	SSE	9	S
10	SSW	11	SW	12	WSW
13	W	14	WNW	15	NW
16	NNW	17	Still wind		

Utilizing web crawler and information extraction techniques, a comprehensive air quality dataset for Hebei Province is built, including 21 variables and 39,622 lines of sample size. The number of missing data in the dataset is less than 5%. The missing data are replaced by the mean values of the preceding and subsequent day's data. As for the abnormal data, they are removed and substituted in the same way. Given the evident dimensional difference among variables (i.e. air quality, atmospheric pollutants data, meteorological data, and spatiotemporal data), all variables are standardized to the range [0, 1] to expedite model convergence speed. It is noteworthy that we integrate latitude and longitude as spatial features of air quality monitoring stations, alongside temporal features such as holidays, seasons, and weeks, as input variables for our model. This enables the fusion of spatiotemporal characteristics and facilitates the joint prediction of air quality across multiple stations. Descriptive statistics of the Hebei Province air quality dataset are presented in Table 3.

V. SPATIOTEMPORAL DISTRIBUTION AND FORECASTING RESULT

In this research, the spatiotemporal distribution of air quality in Hebei Province will be investigated, to ensure its

TABLE 3. Descriptive statistics.

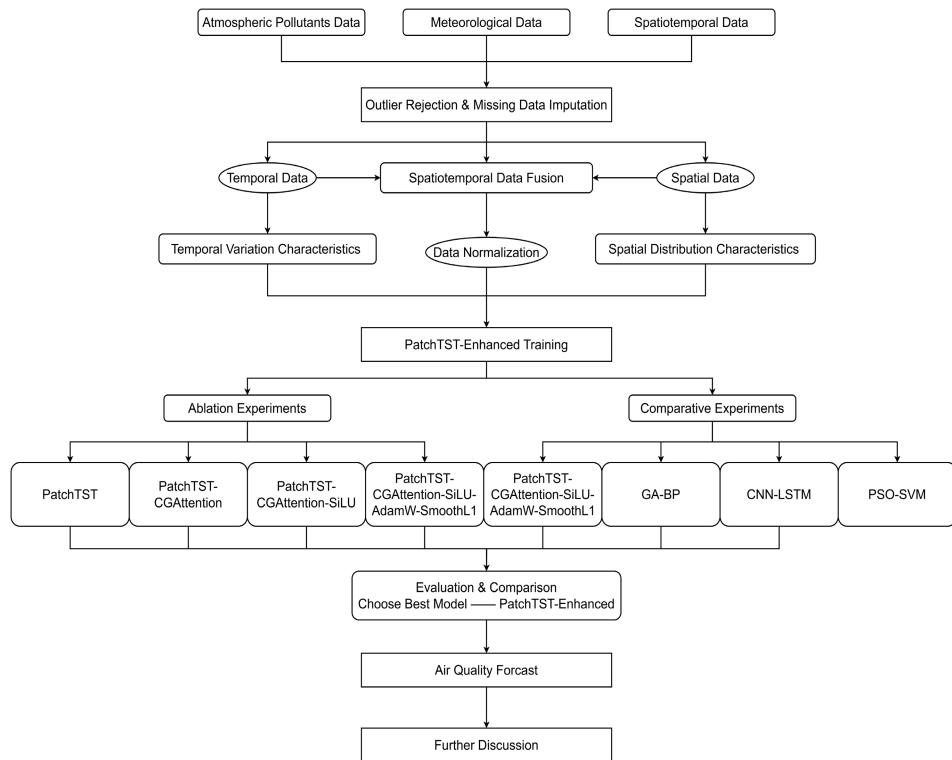
Type	Variable	Unit	Min.	Max.	Mean	Std.	Skewness		Kurtosis		Transformed
							Statistic	SE	Statistic	SE	
Air Quality	AQI	/	9.000	500.000	92.880	62.943	2.340	0.012	7.669	0.025	[0, 1]
Atmospheric Pollutants Data	PM2.5	µg/m³	0.000	796.000	57.760	55.481	2.995	0.012	14.688	0.025	[0, 1]
	PM10	µg/m³	0.000	1971.000	107.300	87.407	3.385	0.012	27.285	0.025	[0, 1]
	SO2	µg/m³	0.000	437.000	23.430	28.623	3.531	0.012	18.473	0.025	[0, 1]
	CO	mg/m³	0.000	18.920	1.083	0.894	3.240	0.012	19.456	0.025	[0, 1]
	NO2	µg/m³	0.000	235.000	38.620	22.016	1.357	0.012	3.165	0.025	[0, 1]
	O3	µg/m³	1.000	296.000	63.980	38.125	0.692	0.012	0.277	0.025	[0, 1]
Meteorological Data	Temperature	°C	-23.000	35.800	13.121	11.406	-0.280	0.012	-1.113	0.025	[0, 1]
	Pressure	hpa	908.300	1044.200	999.508	28.457	-1.456	0.012	1.407	0.025	[0, 1]
	Wind Direction	/	1.000	17.000	8.600	4.567	-0.052	0.012	-1.104	0.025	[0, 1]
	Wind Speed	m/s	0.000	10.800	2.222	1.177	1.286	0.012	2.298	0.025	[0, 1]
	Visibility	km	0.000	35.000	15.377	8.465	0.164	0.012	-1.103	0.025	[0, 1]
	Rain	/	0.000	1.000	0.150	0.361	1.915	0.012	1.667	0.025	[0, 1]
Spatiotemporal Data	Snow	/	0.000	1.000	0.020	0.139	6.920	0.012	45.889	0.025	[0, 1]
	Season	/	1.000	4.000	2.474	1.109	0.037	0.012	-1.338	0.025	[0, 1]
	Quarter	/	1.000	4.000	2.490	1.111	0.011	0.012	-1.343	0.025	[0, 1]
	Latitude	°	36.617	40.967	38.856	1.352	0.007	0.012	-1.102	0.025	[0, 1]
	Longitude	°	114.350	119.517	116.215	1.663	0.536	0.012	-0.904	0.025	[0, 1]
	Week	/	1.000	53.000	26.340	15.015	0.026	0.012	-1.181	0.025	[0, 1]
	Day	/	1.000	7.000	4.000	2.000	0.001	0.012	-1.250	0.025	[0, 1]
	Holiday	/	0.000	1.000	0.320	0.466	0.784	0.012	-1.386	0.025	[0, 1]

significant spatial correlation. Subsequently, the air quality data of Hebei Province, integrated with spatiotemporal features, will be utilized to train the PatchTST-Enhanced model. Based on the model training results, ablation experiments and comparative experiments will be conducted to ultimately determine the optimal model for air quality forecasting in Hebei Province. The detailed experimental scheme is illustrated in Figure 3.

A. SPATIOTEMPORAL DISTRIBUTION CHARACTERISTICS OF AQI IN HEBEI PROVINCE

Utilizing visualization techniques to analyze the seasonal, monthly, and spatial distributions of air quality index (AQI) in Hebei Province. This analysis can facilitate the fusion of spatiotemporal features in the subsequent training process of forecasting models. The results are presented in Figure 4.

From a temporal perspective, the average AQI in Hebei Province during the period from December 2, 2013,

**FIGURE 3.** Flowchart of experimental scheme.

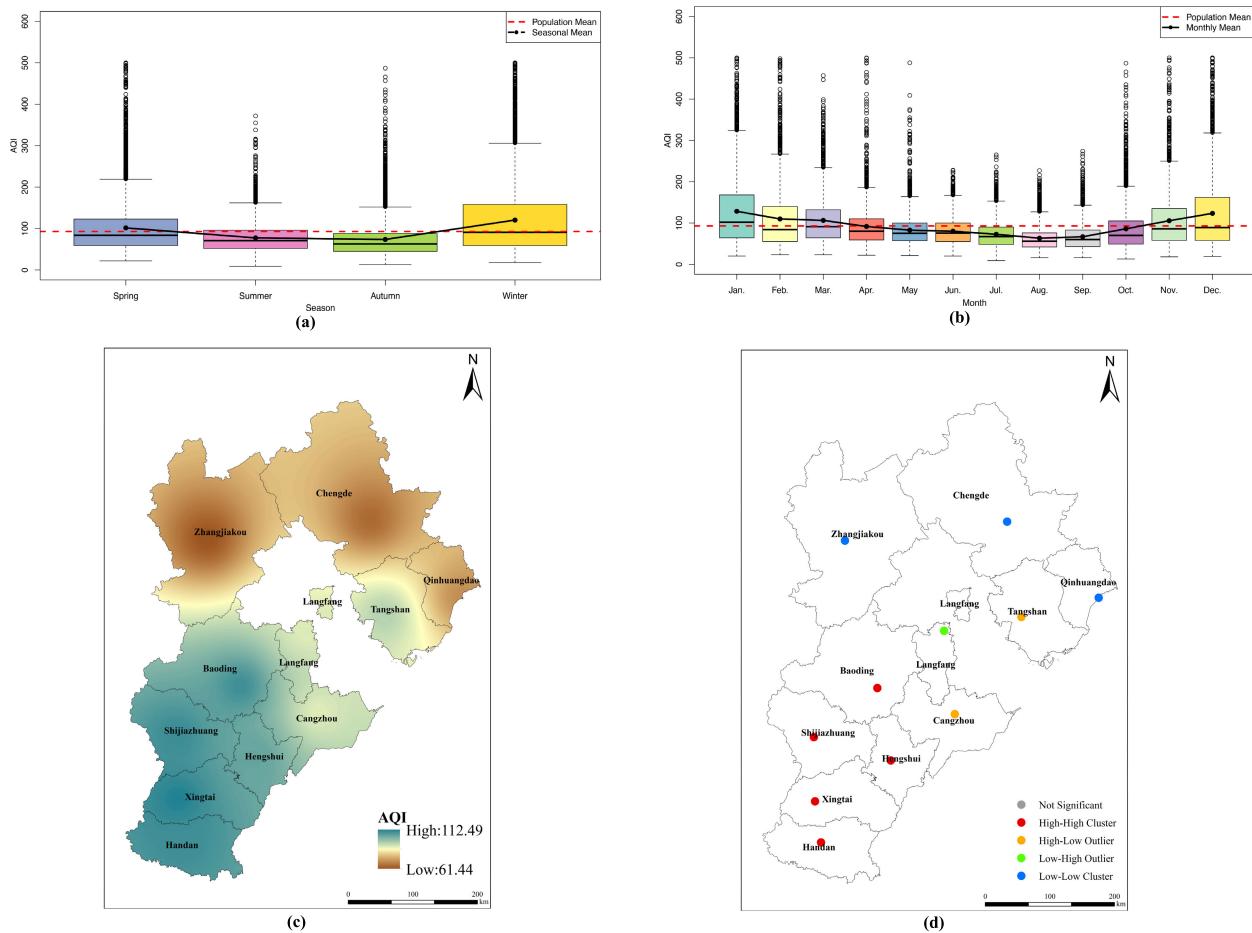


FIGURE 4. Spatiotemporal distribution of AQI in Hebei Province. (a) Seasonal Variation; (b) Monthly Variation; (c) Spatial Distribution; (d) Local Moran's I.

to October 12, 2023, is 92.91. As evident in Figure 4(a), notable seasonal variations in AQI are observed in Hebei Province, displaying a gradual decrease from spring to autumn, followed by a sharp increase in winter. Specifically, the autumn average AQI (73.69 ± 44.436) is substantially lower than the overall average, indicating the optimal air quality during autumn. Conversely, the winter average AQI (120.5 ± 86.216) far exceeds the overall average, suggesting a process of gradual improvement from winter to the subsequent autumn, albeit with severe air pollution during the transition from autumn to winter. Furthermore, air quality exhibits greater stability in autumn, but displays significant variability in winter, leading to a higher probability of extreme weather events [89]. Figure 4 (b) depicts the monthly average AQI in Hebei Province during the same period, ranging from 63.22 to 128.37. Notably, the lowest monthly average AQI occurs in the latter half of the year, forming a right-skewed U-shaped distribution, decreasing from January to August and increasing from August to December. The average AQI from April to October is below the overall average, indicating that 58.33% of the months have better air quality than the annual average. The average AQI in August (63.22 ± 29.818) is the lowest

with remarkable stability, while the average AQI in January (128.37 ± 86.776) is the highest. December also exhibits relatively poor air quality (123.35 ± 94.166) with significant dispersion and instability. The presence of numerous extreme high AQI values in Figure 4 (a) and (b) underscores the frequent occurrence of extremely air quality levels in Hebei Province over the past decade. Thus, incorporating seasonal factors as input variables for prediction is crucial, highlighting the practical significance of this research [90].

From a spatial perspective, the overall Moran's I for AQI across multiple stations in Hebei Province from December 2, 2013, to October 12, 2023, is 0.460841 ($P = 0.002602 < 0.01$), indicating a significant positive spatial correlation and clustering of AQI values [91]. Kriging interpolation of the average AQI across stations within the study period (Figure 4(c)) reveals a southwest-to-northeast gradient of decreasing AQI in Hebei Province [92], [93]. Regions with higher AQI are primarily concentrated in Baoding, Shijiazhuang, Hengshui, Xingtai, and Handan, with a maximum value of 112.52. Conversely, regions with lower AQI are found in Zhangjiakou, Qinhuangdao, Chengde, and Tangshan. The northeastern part of Hebei Province, characterized by higher elevations, exhibits the lowest AQI of

61.46. As depicted in Figure 4(d), the Moran's I of individual stations reveals that five stations in southern Hebei belong to the High-High Cluster. Notably, the provincial capital (Shijiazhuang) and its surrounding areas exhibit the most severe air pollution, with a significant positive spatial correlation in AQI. As the distance from the capital city increases, a declining trend in AQI is observed, indicating an improvement in air quality. Tangshan and Cangzhou are classified as High-Low Outliers, Langfang as Low-High Outliers, while Zhangjiakou, Chengde, and Qinhuangdao constitute the Low-Low Cluster, representing relatively less polluted areas. Based on the above analysis, the existence of significant spatial correlations among these stations underscores the cruciality of incorporating spatiotemporal feature fusion in the process of air quality forecasting.

B. RESULTS OF AQI IN HEBEI PROVINCE

1) RESULT OF PATCHTST-ENHANCED MODEL

In this section, we conduct experiments utilizing the PatchTST-Enhanced model, with prediction lengths (PreLen) set to 96, 192, 336, and 760, respectively, based on previous research. In the PatchTST-Enhanced model, the patches are configured as non-overlapping, and our dataset is subsequently partitioned into 25, 25, 24, and 23 blocks corresponding to the four different prediction lengths. Then, the trained model is employed to forecast the air quality in Hebei Province from December 2, 2013, to October 12, 2023. Figure 5. presents the prediction results of the PatchTST-Enhanced model, and Table 4. shows the Loss Function of the model, which serves as the primary metric for evaluating its performance.

TABLE 4. Loss function of PatchTST-Enhanced.

PreLen	MSE	MAE	RSE
96	0.5408	0.5408	0.5408
192	0.2795	0.2795	0.2795
336	0.6779	0.6779	0.6779
720	0.5763	0.5763	0.5763

According to Figure 5., the Ground Truth curve closely aligns with the Prediction curve, with a high degree of overlap between the blue and orange curves in the first half of each patch, while the overlap diminishes towards the end while maintaining the overall prediction trend. Comparing (a), (b), (c), and (d) in Figure 5., it is observed that as the prediction lengths increase, the area of complete overlap between the blue and orange curves within each patch decreases. From Table 4., the MSE values are concentrated within the range of 0.25-0.7, indicating the superior fitting performance of the PatchTST-Enhanced model. Similarly, MAE and RSE also reside within the desirable range of 0.25-0.7, suggesting proximity between the prediction results and Ground Truth, showing the high prediction accuracy of the PatchTST-Enhanced model. Based on the findings, it can be concluded that the proposed PatchTST-Enhanced model exhibits good performance in multi-scale spatiotemporal feature fusion for

time series prediction tasks and is suitable for air quality forecasting at various stations in Hebei Province.

2) ABLATION EXPERIMENTS

To further validate the effectiveness of the optimization modules in the PatchTST-Enhanced Model, the ablation experiments are conducted. During the parameter tuning for these experiments, the overall structure of each model (e.g., number of layers, hidden layer size) is kept constant, while adjustments are made to the attention mechanism, activation function, optimizer, and loss function. This section utilizes the air quality dataset from Hebei Province to perform experiments under four different Prediction Lengths (PreLen) using PatchTST, PatchTST-CGAttention, PatchTST-CGAttention-SiLU, and PatchTST-CGAttention-SiLU-AdamW-Smooth L1, totaling 16 experiments. The ablation experiment results are presented in Table 5 and Figure 6.

TABLE 5. Result of ablation experiments.

PreLen	CGAttention	SiLU	AdamW	SmoothL1	MSE	MAE	RSE
96	x	x	x	x	0.5835	0.3203	0.7042
	✓	x	x	x	0.5738	0.3202	0.6983
	✓	✓	x	x	0.5623	0.3122	0.6912
	✓	✓	✓	✓	0.5408	0.2795	0.6779
192	x	x	x	x	0.6113	0.3325	0.7191
	✓	x	x	x	0.6021	0.3311	0.7136
	✓	✓	x	x	0.5872	0.3228	0.7047
	✓	✓	✓	✓	0.5763	0.2935	0.6981
336	x	x	x	x	0.6468	0.3547	0.7369
	✓	x	x	x	0.6403	0.3472	0.7331
	✓	✓	x	x	0.6182	0.3383	0.7203
	✓	✓	✓	✓	0.6146	0.3102	0.7182
720	x	x	x	x	0.7694	0.4138	0.7962
	✓	x	x	x	0.7645	0.4078	0.7939
	✓	✓	x	x	0.7498	0.3944	0.7856
	✓	✓	✓	✓	0.7415	0.3680	0.7818

* “x” indicates “not used”, “✓” indicates “used”. The bold data indicate the optimal results.

As shown in Table 5 and Figure 6., the baseline PatchTST model, without any additional modules, has already achieved satisfactory prediction performance for air quality forecasting in Hebei Province, demonstrating its high applicability in deep learning prediction with high-dimensional input. Upon replacing the attention mechanism in PatchTST with CGAttention, a notable improvement in prediction performance can be observed. This enhancement stems primarily from CGAttention's ability to cascade multiple attention layers, thereby strengthening the model's capacity to capture long-distance dependency, leading to enhanced extraction and learning of both local and global features [94], [95]. Further enhancement is achieved by incorporating SiLU into the PatchTST-CGAttention model, which is attributed to SiLU's superior performance in approximating complex functions compared to conventional activation functions like ReLU. Consequently, the PatchTST-CGAttention model with SiLU activation better fits the intricate distributions of temporal data, enhancing its expressive power and generalization ability [78], [96]. By simultaneously employing the AdamW optimizer, SmoothL1 Loss function, and



FIGURE 5. Sample channel signal prediction results of PatchTST-Enhanced Model. (a) PreLen = 96; (b) PreLen = 192; This figure shows the data reconstruction effect of PatchTST-Enhanced Model, with axes unlabeled for units.

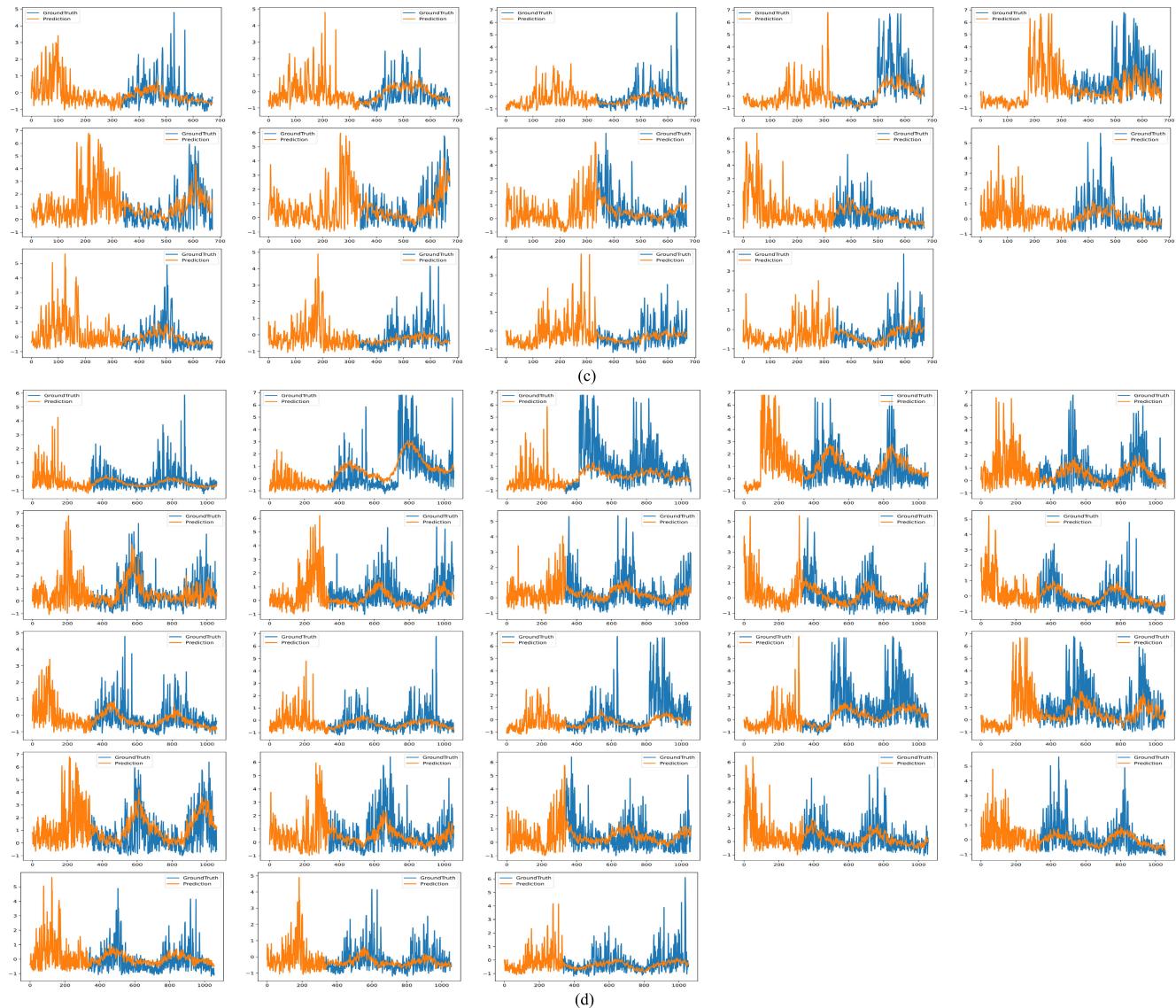


FIGURE 5. (Continued.) Sample channel signal prediction results of PatchTST-Enhanced Model. (c) PreLen = 336; (d) PreLen = 720. This figure shows the data reconstruction effect of PatchTST-Enhanced Model, with axes unlabeled for units.

PatchTST-CGAttention-SiLU, the final PatchTST-Enhanced model is developed. This model achieves performance improvement compared to its predecessors with optimized modules. The primary reasons are twofold: firstly, compared with other complex optimizers, AdamW simplifies implementation and hyperparameter tuning by requiring only learning rate and weight decay adjustments [77], [97]. Secondly, SmoothL1 Loss is easy to implement, typically requiring no additional hyperparameter tuning, and facilitates faster convergence during the optimization process of PatchTST-Enhanced [50]. Based on the above analysis, it can be concluded that the simultaneous utilization of these optimization modules in the PatchTST model effectively enhances its capability for multi-site air quality forecasting.

3) COMPARATIVE EXPERIMENTS

To evaluate the effectiveness of the PatchTST-Enhanced model in integrating spatiotemporal features for air quality forecasting, we conduct a comparative analysis with three prediction models: GA-BP, CNN-LSTM, and PSO-SVM. The rationale for selecting these models lies in two aspects: firstly, they were widely adopted by scholars in the field of deep learning [88], [98], [99], [100]. Secondly, all three models embody continuous optimization and improvements over traditional algorithms, thereby demonstrating robust foundations in time series forecasting. The comparative experiments' results and the corresponding parameters are presented in Figure 7. and Table 6, which offer a comprehensive insight into the performance of the PatchTST-Enhanced model across various dimensions,

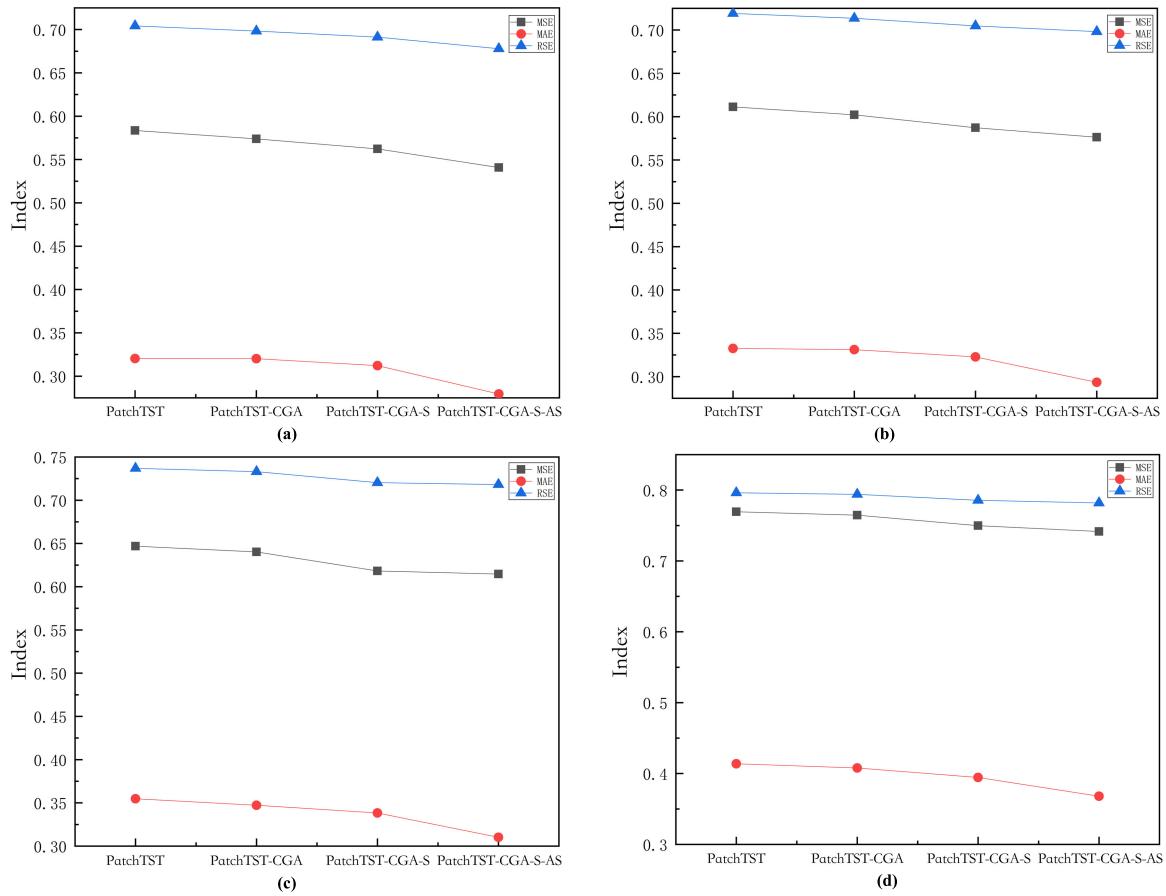


FIGURE 6. Forecasting Performance on Different PreLen. (a) PreLen = 96; (b) PreLen = 192; (c) PreLen = 336; (d) PreLen = 720.

substantiating its rationale for application in air quality forecasting within Hebei Province.

TABLE 6. Result of comparative experiments.

PreLen	Experimental Model	MSE	MAE	RSE
96	PatchTST-Enhanced	0.5408	0.2795	0.6779
	GA-BP	2.7037	1.8720	2.9856
	CNN-LSTM	6.4020	5.3938	6.8789
	PSO-SVM	3.6143	1.8478	3.9759
192	PatchTST-Enhanced	0.6113	0.3325	0.7191
	GA-BP	4.0516	2.9016	4.7689
	CNN-LSTM	6.9471	4.8501	7.0375
	PSO-SVM	4.4676	2.979	5.0357
336	PatchTST-Enhanced	0.6468	0.3547	0.7369
	GA-BP	3.9790	3.0339	4.0375
	CNN-LSTM	8.8426	4.4695	8.9364
	PSO-SVM	5.2640	3.2604	6.0374
720	PatchTST-Enhanced	0.7694	0.4138	0.7962
	GA-BP	10.2010	4.7194	10.5673
	CNN-LSTM	11.2775	5.5528	11.3629
	PSO-SVM	6.7541	3.0283	7.2846

* The bold data indicate the optimal results.

According to Table 6 and Figure 7., it can be seen that the overall prediction of the PatchTST-Enhanced model proposed in this study is better than the other contrast models. Specifically, the PatchTST-Enhanced model significantly outperforms GA-BP, CNN-LSTM, and PSO-SVM in all three metrics, namely, MSE, MAE, and RSE. This may be influenced by the multi-dimensional components

of the air quality data (e.g. time, space, meteorology, and climate) as well as the complex and computationally intensive input variables. After optimizing its attention mechanism, the PatchTST-Enhanced model exhibits a superior capability in handling long sequences and multi-dimensional features. Furthermore, conventional optimization schemes may encounter limitations in seeking optimal parameters, ultimately yielding suboptimal model performance compared to PatchTST-Enhanced. Although a slight decrease in prediction accuracy is observed as the predictive time length extends, these results adequately demonstrate the remarkable predictive advantages of the proposed PatchTST-Enhanced model over existing models in the context of air quality forecasting, particularly in the fusion of spatiotemporal features.

C. FURTHER DISCUSSION

The proposed PatchTST-Enhanced model in this paper has achieved the fusion of spatiotemporal features and long sequence time series modeling under complex input variables. Due to substantial computational demands, all model calculations in this segment are executed using an RTX 4090D 24 GB GPU. Furthermore, given the variety of input variables, there is a risk of overfitting during model training, which is effectively mitigated by using the optimizers and loss functions. Lastly, although the model training in this

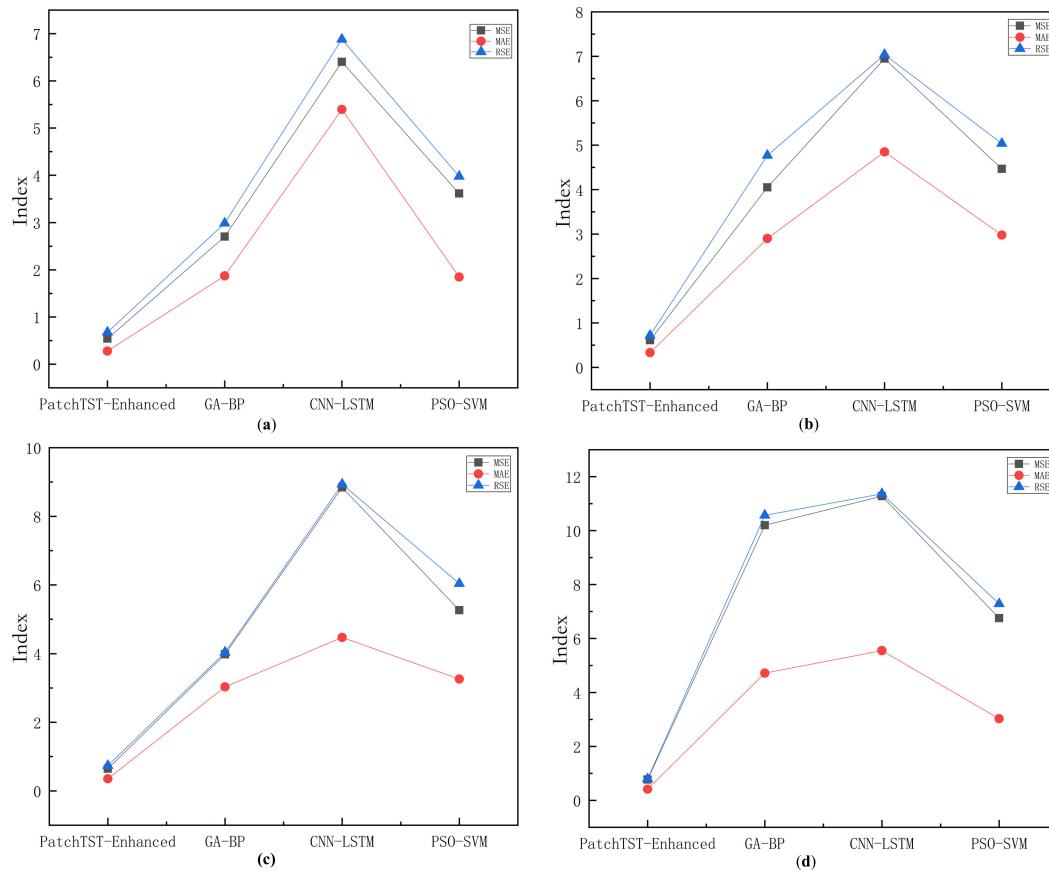


FIGURE 7. Forecasting Performance According to Different Model. (a) PreLen = 96; (b) PreLen = 192; (c) PreLen = 336; (d) PreLen = 720.

paper is exemplified using Hebei Province, China, as a case study, with the longitude and latitude of monitoring stations serving as critical input features, the PatchTST-Enhanced model trained herein can potentially be adapted for air quality prediction in other cities or regions by substituting the input features of longitude and latitude in subsequent research endeavors.

Further research may be conducted from the following two aspects: firstly, future studies may try to construct a multimodal model based on the PatchTST-Enhanced model and to explore the integration of this model with others (e.g., YOLOv5, CNNs) to achieve superior multimodal feature extraction and fusion. By introducing deep learning recognition algorithms into air quality forecasting, more factors or variables related to air quality forecasting can be incorporated into the prediction training process [101], [102], [103]. Secondly, future research may try to introduce adversarial training into the PatchTST-Enhanced model to improve its robustness. It is also suggested to investigate how to implement adaptive learning, enabling the model to automatically adapt to data from diverse tasks and domains [104], [105]. Finally, the PatchTST-Enhanced model holds the potential for real-time air quality monitoring and forecasting in cities or specific regions. Relevant institutions and government departments can leverage these predictions to formulate environmental management policies, plan urban

development, adjust traffic flows, and ultimately improve the environment.

The results of air quality forecasting in Hebei Province through the PatchTST-Enhanced model can be applied in diverse domains. Firstly, precise and long-term air quality forecasting can provide more accurate and useful information for individuals and healthcare institutions, enabling them to take appropriate protective measures, particularly for vulnerable populations such as children, the elderly, and those with respiratory illnesses [4], [106]. Secondly, the findings of air quality forecasting can be applied to the construction of intelligent systems in smart cities, which can provide personalized air quality information and advice to city dwellers [107]. Lastly, air quality forecasting can help to make numerous decisions for both individuals and enterprises. They can guide individuals in selecting suitable times and locations for outdoor activities, improving travel and tourism experiences [108]. For businesses and farmers, it can help them to optimize the management of pollutant emissions, mitigate environmental pollution, and enhance sustainable development in industries and agriculture.

VI. CONCLUSION

In brief, this study endeavors to optimize the PatchTST model and proposes an innovative PatchTST-Enhanced model for air quality forecasting in spatiotemporal fusion. The new

model utilizes atmospheric, pollutants, meteorological, and spatiotemporal data from 11 prefecture-level cities in Hebei Province (from December 2, 2013, to October 12, 2023) as input variables, and then is trained to predict the Air Quality Index (AQI). The performance of the PatchTST-Enhanced model demonstrates robust model capability and predictive accuracy.

Regarding the spatiotemporal distribution, the air quality in Hebei Province exhibits distinct temporal variability and spatial correlation. Notably, the air quality undergoes significant seasonal changes and frequently encounters extreme deterioration. Overall, it follows a gradual improvement trend from spring to autumn, with a sharp deterioration in winter. The Moran's I of the AQI from each station in Hebei Province is 0.460841 ($P = 0.002602 < 0.01$), indicating a significant positive spatial correlation and clustering among the AQI at these stations. This underscores the necessity of incorporating spatiotemporal features into the predictive process of air quality.

According to the training results, the PatchTST-Enhanced model demonstrates robust performance across four distinct prediction lengths (96, 192, 336, 720). Under various patch configurations, over 50% of the predicted values exhibit a high degree of similarity to the Ground Truth, with all predictions aligning perfectly with the Ground Truth trends of increase or decrease. In ablation experiments, the PatchTST-Enhanced model outperforms all optimized combinations of alternative PatchTST models, justifying the rationality of the four enhancement modules designed for the PatchTST model in this study. Comparative experiments reveal that PatchTST-Enhanced significantly improves prediction accuracy and model performance compared to traditional algorithms. This not only underscores the rationale of using PatchTST-Enhanced for multi-site air quality prediction with spatiotemporal feature fusion but also highlights the superiority of the proposed optimization modules over conventional optimization techniques.

In summary, this paper proposes a deep learning method for air quality forecasting under multi-site spatiotemporal feature fusion, which demonstrates excellent performance in forecasting air quality in Hebei Province. Based on our findings, we put forward potential future research directions for the PatchTST-Enhanced model from two perspectives: multi-modal feature extraction and fusion, as well as adversarial training. Lastly, we analyze potential applications of the results of air quality forecasting with PatchTST-Enhanced model.

AUTHOR CONTRIBUTIONS

Conceptualization: Wenyi Cao and Rufei Zhang; methodology: Wenyi Cao; software: Wenyi Cao; validation: Wenyi Cao; formal analysis: Rufei Zhang and Wenxin Cao; investigation: Wenyi Cao and Wenxin Cao; resources: Wenyi Cao and Rufei Zhang; data curation: Wenyi Cao; writing—original draft preparation: Wenyi Cao; writing—review and editing: Rufei Zhang and Wenyi Cao; visualization: Wenyi

Cao; supervision: Rufei Zhang; project administration: Rufei Zhang. All authors have read and agreed to the published version of the manuscript.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGMENT

The authors would like to thank the funders of this project Hebei GEO University, Hebei Center for Ecological and Environmental Geology Research of Hebei GEO University, Southwest Jiaotong University, and all the teams and individuals who supported this research.

REFERENCES

- [1] Y. Kim and V. Radoias, "Severe air pollution exposure and long-term health outcomes," *Int. J. Environ. Res. Public Health*, vol. 19, no. 21, p. 14019, Oct. 2022, doi: [10.3390/ijerph192114019](https://doi.org/10.3390/ijerph192114019).
- [2] C. K. Ward-Caviness and W. E. Cascio, "A narrative review on the impact of air pollution on heart failure risk and exacerbation," *Can. J. Cardiol.*, vol. 39, no. 9, pp. 1244–1252, Sep. 2023, doi: [10.1016/j.cjca.2023.06.423](https://doi.org/10.1016/j.cjca.2023.06.423).
- [3] M. M. Serafini, A. Maddalon, M. Iulini, and V. Galbiati, "Air pollution: Possible interaction between the immune and nervous system?" *Int. J. Environ. Res. Public Health*, vol. 19, no. 23, p. 16037, Nov. 2022, doi: [10.3390/ijerph192316037](https://doi.org/10.3390/ijerph192316037).
- [4] J. Zhu and C. Lu, "Air quality, pollution perception, and residents' health: Evidence from China," *Toxics*, vol. 11, no. 7, p. 591, Jul. 2023, doi: [10.3390/toxics11070591](https://doi.org/10.3390/toxics11070591).
- [5] M. H. Ryu, S. Murphy, M. Hinkley, and C. Carlsten, "COPD exposed to air pollution," *CHEST*, vol. 165, no. 4, pp. 836–846, Apr. 2024, doi: [10.1016/j.chest.2023.11.012](https://doi.org/10.1016/j.chest.2023.11.012).
- [6] D. Dutta and S. K. Pal, "Z-number-based AQI in rough set theoretic framework for interpretation of air quality for different thresholds of PM_{2.5} and PM₁₀," *Environ. Monitor. Assessment*, vol. 194, no. 9, pp. 1–26, Sep. 2022, doi: [10.1007/s10661-022-10325-z](https://doi.org/10.1007/s10661-022-10325-z).
- [7] Y. Chen, D. Li, H. Karimian, S. Wang, and S. Fang, "The relationship between air quality and MODIS aerosol optical depth in major cities of the Yangtze River Delta," *Chemosphere*, vol. 308, Dec. 2022, Art. no. 136301, doi: [10.1016/j.chemosphere.2022.136301](https://doi.org/10.1016/j.chemosphere.2022.136301).
- [8] S. Wang, Y. Ren, and B. Xia, "Estimation of urban AQI based on interpretable machine learning," *Environ. Sci. Pollut. Res.*, vol. 30, no. 42, pp. 96562–96574, Aug. 2023, doi: [10.1007/s11356-023-29336-5](https://doi.org/10.1007/s11356-023-29336-5).
- [9] J. C. M. Pires, S. I. V. Sousa, M. C. Pereira, M. C. M. Alvim-Ferraz, and F. G. Martins, "Management of air quality monitoring using principal component and cluster analysis—Part I: SO₂ and PM₁₀," *Atmos. Environ.*, vol. 42, no. 6, pp. 1249–1260, Feb. 2008, doi: [10.1016/j.atmosenv.2007.10.044](https://doi.org/10.1016/j.atmosenv.2007.10.044).
- [10] R. Yan, J. Liao, J. Yang, W. Sun, M. Nong, and F. Li, "Multi-hour and multi-site air quality index forecasting in Beijing using CNN, LSTM, CNN-LSTM, and spatiotemporal clustering," *Expert Syst. Appl.*, vol. 169, May 2021, Art. no. 114513, doi: [10.1016/j.eswa.2020.114513](https://doi.org/10.1016/j.eswa.2020.114513).
- [11] H. Zhang, Y. Wang, J. Hu, Q. Ying, and X.-M. Hu, "Relationships between meteorological parameters and criteria air pollutants in three megacities in China," *Environ. Res.*, vol. 140, pp. 242–254, Jul. 2015, doi: [10.1016/j.enres.2015.04.004](https://doi.org/10.1016/j.enres.2015.04.004).
- [12] B. Sánchez, M. Sánchez-Muñoz, M. Muñoz-Vicente, G. Cobas, R. Portela, S. Suárez, A. E. González, N. Rodríguez, and R. Amils, "Photocatalytic elimination of indoor air biological and chemical pollution in realistic conditions," *Chemosphere*, vol. 87, no. 6, pp. 625–630, May 2012, doi: [10.1016/j.chemosphere.2012.01.050](https://doi.org/10.1016/j.chemosphere.2012.01.050).
- [13] B. Pandey, M. Agrawal, and S. Singh, "Assessment of air pollution around coal mining area: Emphasizing on spatial distributions, seasonal variations and heavy metals, using cluster and principal component analysis," *Atmos. Pollut. Res.*, vol. 5, no. 1, pp. 79–86, Jan. 2014, doi: [10.5094/apr.2014.010](https://doi.org/10.5094/apr.2014.010).

- [14] J. Zhang, L.-Y. Zhang, M. Du, W. Zhang, X. Huang, Y.-Q. Zhang, Y.-Y. Yang, J.-M. Zhang, S.-H. Deng, F. Shen, Y.-W. Li, and H. Xiao, "Identifying the major air pollutants base on factor and cluster analysis, a case study in 74 Chinese cities," *Atmos. Environ.*, vol. 144, pp. 37–46, Nov. 2016, doi: [10.1016/j.atmosenv.2016.08.066](https://doi.org/10.1016/j.atmosenv.2016.08.066).
- [15] A. Kumar and P. Goyal, "Forecasting of daily air quality index in Delhi," *Sci. Total Environ.*, vol. 409, no. 24, pp. 5517–5523, Nov. 2011, doi: [10.1016/j.scitotenv.2011.08.069](https://doi.org/10.1016/j.scitotenv.2011.08.069).
- [16] S. G. Gocheva-Ilieva, D. S. Voynikova, M. P. Stoimenova, A. V. Ivanov, and I. P. Iliev, "Regression trees modeling of time series for air pollution analysis and forecasting," *Neural Comput. Appl.*, vol. 31, no. 12, pp. 9023–9039, Dec. 2019, doi: [10.1007/s00521-019-04432-1](https://doi.org/10.1007/s00521-019-04432-1).
- [17] A. Gao, J. Wang, J. Poetzscher, S. Li, B. Gao, P. Wang, J. Luo, X. Fang, J. Li, J. Hu, J. Gao, and H. Zhang, "Coordinated health effects attributable to particulate matter and other pollutants exposures in the north China plain," *Environ. Res.*, vol. 208, May 2022, Art. no. 112671, doi: [10.1016/j.envres.2021.112671](https://doi.org/10.1016/j.envres.2021.112671).
- [18] L. Cheng, W. Wei, S. Cheng, C. Zhang, Z. Ye, K. Wang, and R. Wang, "Reductions of multiple air pollutants from coking industry through technology improvements and their impacts on air quality and health risks in a highly industrialized region of China," *Sci. Total Environ.*, vol. 908, Jan. 2024, Art. no. 168360, doi: [10.1016/j.scitotenv.2023.168360](https://doi.org/10.1016/j.scitotenv.2023.168360).
- [19] Q. Liu, M. Zhang, P. Wang, K. Chen, P. Wang, Y. Zhang, B. Zhang, and H. Zhang, "Regional source contributions to fine particulate matter of less studied cities in Beijing-Tianjin-Hebei region in 2017," *Particuology*, vol. 82, pp. 111–121, Nov. 2023, doi: [10.1016/j.partic.2023.01.011](https://doi.org/10.1016/j.partic.2023.01.011).
- [20] S. Gu, S. Li, S. Wu, B. Tian, Y. Hu, M. Cui, and M. Sun, "Study on the spatiotemporal evolution of the ecological landscape and construction of an ecological network: A case study of Hebei province," *Sustainability*, vol. 15, no. 21, p. 15661, Nov. 2023, doi: [10.3390/su152115661](https://doi.org/10.3390/su152115661).
- [21] Q. Xing, Z. Yang, and W. Yao, "Air quality assessment based on R/SPA composite model: A case study of the Hebei province (China)," *Frontiers Public Health*, vol. 10, Oct. 2022, Art. no. 922125, doi: [10.3389/fpubh.2022.922125](https://doi.org/10.3389/fpubh.2022.922125).
- [22] M. Emeç and M. Yurtsever, "A novel ensemble machine learning method for accurate air quality prediction," *Int. J. Environ. Sci. Technol.*, pp. 1–18, May 2024, doi: [10.1007/s13762-024-05671-z](https://doi.org/10.1007/s13762-024-05671-z).
- [23] S. Mahmud, T. B. I. Ridi, M. S. Miah, F. Sarower, and S. Elahee, "Implementing machine learning algorithms to predict particulate matter ($PM_{2.5}$): A case study in the Paso del Norte region," *Atmosphere*, vol. 13, no. 12, p. 2100, Dec. 2022, doi: [10.3390/atmos13122100](https://doi.org/10.3390/atmos13122100).
- [24] Z. Wu, Y. Tian, M. Li, B. Wang, Y. Quan, and J. Liu, "Prediction of air pollutant concentrations based on the long short-term memory neural network," *J. Hazardous Mater.*, vol. 465, Mar. 2024, Art. no. 133099, doi: [10.1016/j.jhazmat.2023.133099](https://doi.org/10.1016/j.jhazmat.2023.133099).
- [25] Y. Cao, D. Zhang, S. Ding, W. Zhong, and C. Yan, "A hybrid air quality prediction model based on empirical mode decomposition," *Tsinghua Sci. Technol.*, vol. 29, no. 1, pp. 99–111, Feb. 2024, doi: [10.26599/tst.2022.9010060](https://doi.org/10.26599/tst.2022.9010060).
- [26] B. Houria, M. Abderrahmane, K. Kenza, and G. Gábor, "Short-term predictions of PM_{10} and NO_2 concentrations in urban environments based on ARIMA search grid modeling," *CLEAN Soil, Air, Water*, vol. 52, no. 6, Jun. 2024, Art. no. 2300395, doi: [10.1002/clean.202300395](https://doi.org/10.1002/clean.202300395).
- [27] Y. Guo, T. Zhu, Z. Li, and C. Ni, "Auto-modal: Air-quality index forecasting with modal decomposition attention," *Sensors*, vol. 22, no. 18, p. 6953, Sep. 2022, doi: [10.3390/s22186953](https://doi.org/10.3390/s22186953).
- [28] M. Wang, Y. Zhang, J. C. H. Fung, C. Lin, and A. K. H. Lau, "A novel hybrid clustering model of region segmentation to fuse CMAQ simulations with observations," *Atmos. Environ.*, vol. 278, Jun. 2022, Art. no. 119062, doi: [10.1016/j.atmosenv.2022.119062](https://doi.org/10.1016/j.atmosenv.2022.119062).
- [29] Y. Ji, D. Wang, Q. Li, T. Liu, and Y. Bai, "Global wildfire danger predictions based on deep learning taking into account static and dynamic variables," *Forests*, vol. 15, no. 1, p. 216, Jan. 2024, doi: [10.3390/f15010216](https://doi.org/10.3390/f15010216).
- [30] R. Luo, Y. Song, L. Huang, Y. Zhang, and R. Su, "AST-GIN: Attribute-augmented spatiotemporal graph informer network for electric vehicle charging station availability forecasting," *Sensors*, vol. 23, no. 4, p. 1975, Feb. 2023, doi: [10.3390/s23041975](https://doi.org/10.3390/s23041975).
- [31] Z. Wu, J. He, S. Hu, and J. Wen, "The prediction of two-dimensional intelligent ocean temperature based on deep learning," *Expert Syst.*, p. 13367, May 2023, doi: [10.1111/exsy.13367](https://doi.org/10.1111/exsy.13367).
- [32] M. Asgari, W. Yang, and M. Farnaghi, "Spatiotemporal data partitioning for distributed random forest algorithm: Air quality prediction using imbalanced big spatiotemporal data on spark distributed framework," *Environ. Technol. Innov.*, vol. 27, Aug. 2022, Art. no. 102776, doi: [10.1016/j.eti.2022.102776](https://doi.org/10.1016/j.eti.2022.102776).
- [33] H. Liu, Q. Han, H. Sun, J. Sheng, and Z. Yang, "Spatiotemporal adaptive attention graph convolution network for city-level air quality prediction," *Sci. Rep.*, vol. 13, no. 1, pp. 1–16, Aug. 2023, doi: [10.1038/s41598-023-39286-0](https://doi.org/10.1038/s41598-023-39286-0).
- [34] S. Sui and Q. Han, "Multi-view multi-task spatiotemporal graph convolutional network for air quality prediction," *Sci. Total Environ.*, vol. 893, Oct. 2023, Art. no. 164699, doi: [10.1016/j.scitotenv.2023.164699](https://doi.org/10.1016/j.scitotenv.2023.164699).
- [35] Z. Li, X. Zhang, and Z. Dong, "TSF-transformer: A time series forecasting model for exhaust gas emission using transformer," *Appl. Intell.*, vol. 53, no. 13, pp. 17211–17225, Jul. 2023, doi: [10.1007/s10489-022-04326-1](https://doi.org/10.1007/s10489-022-04326-1).
- [36] L. Xiong, L. Su, X. Wang, and C. Pan, "Dynamic adaptive graph convolutional transformer with broad learning system for multi-dimensional chaotic time series prediction," *Appl. Soft Comput.*, vol. 157, May 2024, Art. no. 111516, doi: [10.1016/j.asoc.2024.111516](https://doi.org/10.1016/j.asoc.2024.111516).
- [37] L. Ren, H. Wang, T. Mo, and L. T. Yang, "A lightweight group transformer-based time series reduction network for edge intelligence and its application in industrial RUL prediction," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Jan. 3, 2024, doi: [10.1109/tnnls.2023.3347227](https://doi.org/10.1109/tnnls.2023.3347227).
- [38] Y. Nie, N. H. Nguyen, P. Sinthong, and J. Kalagnanam, "A time series is worth 64 words: Long-term forecasting with transformers," presented at the 11th Int. Conf. Learn. Represent., Kigali Rwanda, May 2023.
- [39] T. Suzuki and Y. Aoki, "Efficient transformer-based compressed video modeling via informative patch selection," *Sensors*, vol. 23, no. 1, p. 244, Dec. 2022, doi: [10.3390/s23010244](https://doi.org/10.3390/s23010244).
- [40] X. Chen, M. Ma, Y. Li, S. Mei, Z. Han, J. Zhao, and W. Cheng, "Hierarchical feature fusion of transformer with patch dilating for remote sensing scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 4410516, doi: [10.1109/TGRS.2023.3331880](https://doi.org/10.1109/TGRS.2023.3331880).
- [41] H. Chen, Z. Wang, H. Tian, L. Yuan, X. Wang, and P. Leng, "A robust visual tracking method based on reconstruction patch transformer tracking," *Sensors*, vol. 22, no. 17, p. 6558, Aug. 2022, doi: [10.3390/s22176558](https://doi.org/10.3390/s22176558).
- [42] X. Liu, H. Peng, N. Zheng, Y. Yang, H. Hu, and Y. Yuan, "EfficientViT: Memory efficient vision transformer with cascaded group attention," presented at the IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Seattle WA, USA, Jun. 2023.
- [43] W. Ye, "Automatic detection and localization of surface defects in high-speed railway ballastless track based on cascaded group attention and optoelectronic encoder," *Available SSRN*. Amsterdam, The Netherlands: Elsevier, Jan. 2024, doi: [10.2139/ssrn.4819840](https://doi.org/10.2139/ssrn.4819840).
- [44] S. K. R. Chinnam, V. Sistla, and V. K. K. Kolli, "Multimodal attention-gated cascaded U-Net model for automatic brain tumor detection and segmentation," *Biomed. Signal Process. Control*, vol. 78, Sep. 2022, Art. no. 103907, doi: [10.1016/j.bspc.2022.103907](https://doi.org/10.1016/j.bspc.2022.103907).
- [45] S. Wang, C. Hou, Y. Chen, Z. Liu, Z. Zhang, and G. Zhang, "Classification of hyperspectral and LiDAR data using multi-modal transformer cascaded fusion net," *Remote Sens.*, vol. 15, no. 17, p. 4142, Aug. 2023, doi: [10.3390/rs15174142](https://doi.org/10.3390/rs15174142).
- [46] X. Qi and X. Dong, "Improved Yolov7-tiny algorithm for steel surface defect detection," *Comput. Eng. Appl.*, vol. 59, pp. 176–183, Jan. 2023, doi: [10.3778/j.issn.1002-8331.2302-0191](https://doi.org/10.3778/j.issn.1002-8331.2302-0191).
- [47] T. Hussain and H. Shouño, "MAGRes-UNet: Improved medical image segmentation through a deep learning paradigm of multi-attention gated residual U-Net," *IEEE Access*, vol. 12, pp. 40290–40310, 2024, doi: [10.1109/ACCESS.2024.3374108](https://doi.org/10.1109/ACCESS.2024.3374108).
- [48] C. Zhou, H. Wang, Y. Liu, X. Ni, and Y. Liu, "Green plums surface defect detection based on deep learning methods," *IEEE Access*, vol. 10, pp. 100397–100407, 2022, doi: [10.1109/ACCESS.2022.3206864](https://doi.org/10.1109/ACCESS.2022.3206864).
- [49] L. Hou, L. Yu, S. Tian, and Y. Zhang, "FMAGAN: Fusing multiple attention and generative adversarial network to enhance underwater image," *J. Intell. Fuzzy Syst.*, vol. 42, no. 3, pp. 2421–2433, Feb. 2022, doi: [10.3233/JIFS-211680](https://doi.org/10.3233/JIFS-211680).
- [50] Z. Zhu, Y. Lei, Y. Qin, C. Zhu, and Y. Zhu, "IRE: Improved image super-resolution based on real-ESRGAN," *IEEE Access*, vol. 11, pp. 45334–45348, 2023, doi: [10.1109/ACCESS.2023.3256086](https://doi.org/10.1109/ACCESS.2023.3256086).
- [51] D. Guo, Z. Cao, H. Fu, and Z. Li, "Remaining useful life estimation for rolling bearings using MSGCNN-TR," *IEEE Sensors J.*, vol. 22, no. 24, pp. 24333–24343, Dec. 2022, doi: [10.1109/JSEN.2022.3221753](https://doi.org/10.1109/JSEN.2022.3221753).

- [52] H. Liu, C. Chen, R. Hu, J. Bin, H. Dong, and Z. Liu, “CGTD-Net: Channel-wise global transformer-based dual-branch network for industrial strip steel surface defect detection,” *IEEE Sensors J.*, vol. 24, no. 4, pp. 4863–4873, Feb. 2024, doi: [10.1109/JSEN.2023.3346470](https://doi.org/10.1109/JSEN.2023.3346470).
- [53] J. Zhang, X. Wu, and T. E. Chow, “Space-time cluster’s detection and geographical weighted regression analysis of COVID-19 mortality on Texas counties,” *Int. J. Environ. Res. Public Health*, vol. 18, no. 11, p. 5541, May 2021, doi: [10.3390/ijerph18115541](https://doi.org/10.3390/ijerph18115541).
- [54] D. Le, Y. Li, and F. Ren, “Does air quality improvement promote enterprise productivity increase? Based on the spatial spillover effect of 242 cities in China,” *Frontiers Public Health*, vol. 10, Nov. 2022, Art. no. 1050971, doi: [10.3389/fpubh.2022.1050971](https://doi.org/10.3389/fpubh.2022.1050971).
- [55] H. Luo, N. Wang, J. Chen, X. Ye, and Y.-F. Sun, “Study on the thermal effects and air quality improvement of green roof,” *Sustainability*, vol. 7, no. 3, pp. 2804–2817, Mar. 2015, doi: [10.3390/su7032804](https://doi.org/10.3390/su7032804).
- [56] N. Jiang, C. Ao, L. Xu, Y. Wei, and Y. Long, “Will information interventions affect public preferences and willingness to pay for air quality improvement? An empirical study based on deliberative choice experiment,” *Sci. Total Environ.*, vol. 868, Apr. 2023, Art. no. 161436, doi: [10.1016/j.scitotenv.2023.161436](https://doi.org/10.1016/j.scitotenv.2023.161436).
- [57] F. N. A. Suris, M. A. A. Bakar, N. M. Ariff, M. S. M. Nadzir, and K. Ibrahim, “Malaysia PM₁₀ air quality time series clustering based on dynamic time warping,” *Atmosphere*, vol. 13, no. 4, p. 503, Mar. 2022, doi: [10.3390/atmos13040503](https://doi.org/10.3390/atmos13040503).
- [58] W. Zhou, Y. Zhuang, and Y. Chen, “Spatial convergence and transfer path of atmospheric emission efficiency: An empirical analysis from China,” *J. Cleaner Prod.*, vol. 414, Aug. 2023, Art. no. 137675, doi: [10.1016/j.jclepro.2023.137675](https://doi.org/10.1016/j.jclepro.2023.137675).
- [59] P. Hajek and V. Olej, “Predicting common air quality index—The case of Czech microregions,” *Aerosol Air Quality Res.*, vol. 15, no. 2, pp. 544–555, 2015, doi: [10.4209/aaqr.2014.08.0154](https://doi.org/10.4209/aaqr.2014.08.0154).
- [60] A. Challoner, F. Pilla, and L. Gill, “Prediction of indoor air exposure from outdoor air quality using an artificial neural network model for inner city commercial buildings,” *Int. J. Environ. Res. Public Health*, vol. 12, no. 12, pp. 15233–15253, Dec. 2015, doi: [10.3390/ijerph121214975](https://doi.org/10.3390/ijerph121214975).
- [61] G. G. Garner and A. M. Thompson, “The value of air quality forecasting in the mid-atlantic region,” *Weather, Climate, Soc.*, vol. 4, no. 1, pp. 69–79, Jan. 2012, doi: [10.1175/wcas-d-10-05010.1](https://doi.org/10.1175/wcas-d-10-05010.1).
- [62] U. Kumar and V. K. Jain, “ARIMA forecasting of ambient air pollutants (O₃, NO, NO₂ and CO),” *Stochastic Environ. Res. Risk Assessment*, vol. 24, no. 5, pp. 751–760, Jul. 2010, doi: [10.1007/s00477-009-0361-8](https://doi.org/10.1007/s00477-009-0361-8).
- [63] W. Sun and J. Sun, “Daily PM_{2.5} concentration prediction based on principal component analysis and LSSVM optimized by cuckoo search algorithm,” *J. Environ. Manage.*, vol. 188, pp. 144–152, Mar. 2017, doi: [10.1016/j.jenvman.2016.12.011](https://doi.org/10.1016/j.jenvman.2016.12.011).
- [64] T. Liu, A. K. H. Lau, K. Sandbrink, and J. C. H. Fung, “Time series forecasting of air quality based on regional numerical modeling in Hong Kong,” *J. Geophys. Res., Atmos.*, vol. 123, no. 8, pp. 4175–4196, Apr. 2018, doi: [10.1002/2017jd028052](https://doi.org/10.1002/2017jd028052).
- [65] P.-P. Xiong, W.-J. Yan, G.-Z. Wang, and L.-L. Pei, “Grey extended prediction model based on IRLS and its application on smog pollution,” *Appl. Soft Comput.*, vol. 80, pp. 797–809, Jul. 2019, doi: [10.1016/j.asoc.2019.04.035](https://doi.org/10.1016/j.asoc.2019.04.035).
- [66] S. Zhu, X. Lian, H. Liu, J. Hu, Y. Wang, and J. Che, “Daily air quality index forecasting with hybrid models: A case in China,” *Environ. Pollut.*, vol. 231, pp. 1232–1244, Dec. 2017, doi: [10.1016/j.envpol.2017.08.069](https://doi.org/10.1016/j.envpol.2017.08.069).
- [67] Y. Li and Y. Tao, “Daily PM₁₀ concentration forecasting based on multiscale fusion support vector regression,” *J. Intell. Fuzzy Syst.*, vol. 34, no. 6, pp. 3833–3844, Jun. 2018, doi: [10.3233/jifs-169555](https://doi.org/10.3233/jifs-169555).
- [68] J. Liu and J. Xing, “Identifying contributors to PM_{2.5} simulation biases of chemical transport model using fully connected neural networks,” *J. Adv. Model. Earth Syst.*, vol. 15, no. 2, Feb. 2023, Art. no. e2021MS002898, doi: [10.1029/2021ms002898](https://doi.org/10.1029/2021ms002898).
- [69] B. Zhang, G. Zou, D. Qin, Q. Ni, H. Mao, and M. Li, “RCL-learning: ResNet and convolutional long short-term memory-based spatiotemporal air pollutant concentration prediction model,” *Expert Syst. Appl.*, vol. 207, Nov. 2022, Art. no. 118017, doi: [10.1016/j.eswa.2022.118017](https://doi.org/10.1016/j.eswa.2022.118017).
- [70] J. Wang, X. Li, J. Li, Q. Sun, and H. Wang, “NGCU: A new RNN model for time-series data prediction,” *Big Data Res.*, vol. 27, Feb. 2022, Art. no. 100296, doi: [10.1016/j.bdr.2021.100296](https://doi.org/10.1016/j.bdr.2021.100296).
- [71] R. Bala and R. P. Singh, “A dual-stage advanced deep learning algorithm for long-term and long-sequence prediction for multivariate financial time series,” *Appl. Soft Comput.*, vol. 126, Sep. 2022, Art. no. 109317, doi: [10.1016/j.asoc.2022.109317](https://doi.org/10.1016/j.asoc.2022.109317).
- [72] G. Yang, Y. Qian, H. Liu, B. Tang, R. Qi, Y. Lu, and J. Geng, “MSFusion: Multistage for remote sensing image spatiotemporal fusion based on texture transformer and convolutional neural network,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 4653–4666, 2022, doi: [10.1109/JSTARS.2022.3179415](https://doi.org/10.1109/JSTARS.2022.3179415).
- [73] E. Alerskans, J. Nyborg, M. Birk, and E. Kaas, “A transformer neural network for predicting near-surface temperature,” *Meteorolog. Appl.*, vol. 29, no. 5, Sep. 2022, Art. no. e2098, doi: [10.1002/met.2098](https://doi.org/10.1002/met.2098).
- [74] G. Ren, M. Lazarou, J. Yuan, and T. Stathaki, “Towards automated polyp segmentation using weakly- and semi-supervised learning and deformable transformers,” presented at the IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW), Seattle WA, USA, Jun. 2023.
- [75] H. Jiang, Y. Qian, G. Yang, and H. Liu, “MLKNet: Multi-stage for remote sensing image spatiotemporal fusion network based on a large kernel attention,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 1257–1268, 2024, doi: [10.1109/JSTARS.2023.3338978](https://doi.org/10.1109/JSTARS.2023.3338978).
- [76] J. Liu, C. Gong, S. Chen, and N. Zhou, “Multi-Step-Ahead wind speed forecast method based on outlier correction, optimized decomposition, and DLinear model,” *Mathematics*, vol. 11, no. 12, p. 2746, Jun. 2023, doi: [10.3390/math11122746](https://doi.org/10.3390/math11122746).
- [77] L. Guan, “Weight prediction boosts the convergence of adamw,” presented at the Pacific–Asia Conf. Knowl. Discovery Data Mining, Cham, Switzerland, 2023.
- [78] S. Verma, A. Chug, A. P. Singh, and D. Singh, “PDS-MCNet: A hybrid framework using MobileNetV2 with SiLU6 activation function and capsule networks for disease severity estimation in plants,” *Neural Comput. Appl.*, vol. 35, no. 25, pp. 18641–18664, Sep. 2023, doi: [10.1007/s00521-023-08693-9](https://doi.org/10.1007/s00521-023-08693-9).
- [79] P. Liu, C. Fang, and Z. Qiao, “A dense and U-shaped transformer with dual-domain multi-loss function for sparse-view CT reconstruction,” *J. X-Ray Sci. Technol.*, vol. 32, no. 2, pp. 207–228, Mar. 2024, doi: [10.3233/xst-230184](https://doi.org/10.3233/xst-230184).
- [80] T. Nishiyama, A. Kumagai, K. Kamiya, and K. Takahashi, “SILU: Strategy involving large-scale unlabeled logs for improving malware detector,” presented at the IEEE Symp. Comput. Commun. (ISCC), Rennes, France, Jul. 2020.
- [81] P. Zhou, X. Xie, Z. Lin, K. C. Toh, and S. Yan, “Win: Weight-decay-integrated Nesterov acceleration for faster network training,” *J. Mach. Learn. Res.*, vol. 25, no. 83, pp. 1–74, Jan. 2024.
- [82] J. Lu and X. Li, “Improved gradient-descent-based control algorithms for multi-agent systems with fixed and switching topology,” *Int. J. Control. Autom. Syst.*, vol. 21, no. 11, pp. 3673–3683, Nov. 2023, doi: [10.1007/s12555-022-0513-x](https://doi.org/10.1007/s12555-022-0513-x).
- [83] Y. Li, Z. Jia, Z. Liu, H. Shao, W. Zhao, Z. Liu, and B. Wang, “Interpretable intelligent fault diagnosis strategy for fixed-wing UAV elevator fault diagnosis based on improved cross entropy loss,” *Meas. Sci. Technol.*, vol. 35, no. 7, Jul. 2024, Art. no. 076110, doi: [10.1088/1361-6501/ad3666](https://doi.org/10.1088/1361-6501/ad3666).
- [84] J. Li, Z. Zhang, R. Song, Y. Li, and Q. Du, “SCFormer: Spectral coordinate transformer for cross-domain few-shot hyperspectral image classification,” *IEEE Trans. Image Process.*, vol. 33, pp. 840–855, 2024, doi: [10.1109/TIP.2024.3351443](https://doi.org/10.1109/TIP.2024.3351443).
- [85] E. Seifossadat and H. Sameti, “Data-to-text generation using conditional generative adversarial with enhanced transformer,” *Natural Lang. Eng.*, pp. 1–26, Nov. 2023, doi: [10.1017/s1351324923000487](https://doi.org/10.1017/s1351324923000487).
- [86] J. Yang, Y. Yin, L. Yang, S. Ma, H. Huang, D. Zhang, F. Wei, and Z. Li, “GTrans: Grouping and fusing transformer layers for neural machine translation,” *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 31, pp. 1489–1498, 2023, doi: [10.1109/TASLP.2022.3221040](https://doi.org/10.1109/TASLP.2022.3221040).
- [87] I. Loshchilov and F. Hutter, “Decoupled weight decay regularization,” presented at the 7th Int. Conf. Learn. Represent., May 2019.
- [88] M. Ning, J. Guan, P. Liu, Z. Zhang, and G. M. P. O’Hare, “GA-BP air quality evaluation method based on fuzzy theory,” *Comput. Mater. Continua*, vol. 58, no. 1, pp. 215–227, 2019, doi: [10.32604/cmc.2019.03763](https://doi.org/10.32604/cmc.2019.03763).
- [89] S. Sukkhum, A. Lim, T. Ingviya, and R. Saelim, “Seasonal patterns and trends of air pollution in the upper northern Thailand from 2004 to 2018,” *Aerosol Air Qual. Res.*, vol. 22, no. 5, 2022, Art. no. 210318, doi: [10.4209/aaqr.210318](https://doi.org/10.4209/aaqr.210318).

- [90] N. Das, A. Sagar, R. Bhattacharjee, A. K. Agnihotri, A. Ohri, and S. Gaur, "Time series forecasting of temperature and turbidity due to global warming in river ganga at and around Varanasi, India," *Environ. Monitor. Assessment*, vol. 194, no. 9, pp. 1–27, Sep. 2022, doi: [10.1007/s10661-022-10274-7](https://doi.org/10.1007/s10661-022-10274-7).
- [91] F. Guo, Z. Wang, S. Ji, and Q. Lu, "Influential nodes identification in the air pollution spatial correlation weighted networks and collaborative governance: Taking China's three urban agglomerations as examples," *Int. J. Environ. Res. Public Health*, vol. 19, no. 8, p. 4461, Apr. 2022, doi: [10.3390/ijerph19084461](https://doi.org/10.3390/ijerph19084461).
- [92] C. Xu, J. Wang, M. Hu, and W. Wang, "A new method for interpolation of missing air quality data at monitor stations," *Environ. Int.*, vol. 169, Nov. 2022, Art. no. 107538, doi: [10.1016/j.envint.2022.107538](https://doi.org/10.1016/j.envint.2022.107538).
- [93] K. K. R. S. Samal, K. S. Babu, and S. K. Das, "Spatial-temporal prediction of air quality by deep learning and Kriging interpolation approach," *ICST Trans. Scalable Inf. Syst.*, vol. 10, pp. 1–14, Aug. 2023, doi: [10.4108/eetsis.3325](https://doi.org/10.4108/eetsis.3325).
- [94] T. MeshuWelde and L. Liao, "Counting-based visual question answering with serial cascaded attention deep learning," *Pattern Recognit.*, vol. 144, Dec. 2023, Art. no. 109850, doi: [10.1016/j.patcog.2023.109850](https://doi.org/10.1016/j.patcog.2023.109850).
- [95] Y. Huang, K. Wang, R. Zhuo, X. He, J. Wu, H. Wang, Z. Liang, and X. Wang, "Non-mechanical self-alignment system for free-space optical communications based on a cascaded liquid crystal optical antenna," *Opt. Exp.*, vol. 31, no. 2, p. 929, 2023, doi: [10.1364/oe.477316](https://doi.org/10.1364/oe.477316).
- [96] B. Singh, S. Patel, A. Vijayvargiya, and R. Kumar, "Analyzing the impact of activation functions on the performance of the data-driven gait model," *Results Eng.*, vol. 18, Jun. 2023, Art. no. 101029, doi: [10.1016/j.rineng.2023.101029](https://doi.org/10.1016/j.rineng.2023.101029).
- [97] Z. Lu, X. Li, and P. Qiu, "An AdamW-based deep neural network using feature selection and data oversampling for intrusion detection," in *Proc. 8th Int. Conf. Data Sci. Cyberspace (DSC)*, vol. 34, Hefei, China, Aug. 2023.
- [98] J. Zhang and S. Li, "Air quality index forecast in Beijing based on CNN-LSTM multi-model," *Chemosphere*, vol. 308, Dec. 2022, Art. no. 136180, doi: [10.1016/j.chemosphere.2022.136180](https://doi.org/10.1016/j.chemosphere.2022.136180).
- [99] S. Gunasekar, G. J. R. Kumar, and Y. D. Kumar, "Sustainable optimized LSTM-based intelligent system for air quality prediction in Chennai," *Acta Geophys.*, vol. 70, no. 6, pp. 2889–2899, May 2022, doi: [10.1007/s11600-022-00796-6](https://doi.org/10.1007/s11600-022-00796-6).
- [100] W. Liu, G. Guo, F. Chen, and Y. Chen, "Meteorological pattern analysis assisted daily PM_{2.5} grades prediction using SVM optimized by PSO algorithm," *Atmos. Pollut. Res.*, vol. 10, no. 5, pp. 1482–1491, Sep. 2019, doi: [10.1016/j.apr.2019.04.005](https://doi.org/10.1016/j.apr.2019.04.005).
- [101] Q. Zhao, B. Liu, S. Lyu, C. Wang, and H. Zhang, "TPH-YOLOv5++: Boosting object detection on drone-captured scenarios with cross-layer asymmetric transformer," *Remote Sens.*, vol. 15, no. 6, p. 1687, Mar. 2023, doi: [10.3390/rs15061687](https://doi.org/10.3390/rs15061687).
- [102] M. Lv and W.-H. Su, "YOLOv5-CBAM-C3TR: An optimized model based on transformer module and attention mechanism for apple leaf disease detection," *Frontiers Plant Sci.*, vol. 14, Jan. 2024, Art. no. 1323301, doi: [10.3389/fpls.2023.1323301](https://doi.org/10.3389/fpls.2023.1323301).
- [103] W. Cao, P. Lu, and W. Cao, "Multimodal gesture recognition with spatio-temporal features fusion based on YOLOv5 and MediaPipe," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 38, no. 8, Jun. 2024, Art. no. 2455007, doi: [10.1142/S0218001424550073](https://doi.org/10.1142/S0218001424550073).
- [104] Y. Zheng and D. Wang, "An auxiliary classifier generative adversarial network based fault diagnosis for analog circuit," *IEEE Access*, vol. 11, pp. 86824–86833, 2023, doi: [10.1109/ACCESS.2023.3305261](https://doi.org/10.1109/ACCESS.2023.3305261).
- [105] S. Hao, Y. Xia, and Y. Ye, "Generative adversarial network with transformer for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, pp. 1–5, 2023, doi: [10.1109/LGRS.2023.3322139](https://doi.org/10.1109/LGRS.2023.3322139).
- [106] S. K. Thakrar, J. A. Johnson, and S. Polasky, "Land-use decisions have substantial air quality health effects," *Environ. Sci. Technol.*, vol. 58, no. 1, pp. 381–390, Jan. 2024, doi: [10.1021/acs.est.3c02280](https://doi.org/10.1021/acs.est.3c02280).
- [107] N. K. Kim, D. H. Kang, B. W. Kim, and H. W. Kang, "Optimal location and performance prediction of portable air cleaner in composite room shapes using convolutional neural network," *Building Environ.*, vol. 242, Aug. 2023, Art. no. 110500, doi: [10.1016/j.buildenv.2023.110500](https://doi.org/10.1016/j.buildenv.2023.110500).
- [108] H. He and Z. Li, "Air quality and residents' health in China: An empirical analysis based on spatial simultaneous equations," *Appl. Sci.*, vol. 12, no. 23, p. 11897, Nov. 2022, doi: [10.3390/app122311897](https://doi.org/10.3390/app122311897).



WENYI CAO is currently pursuing the master's degree in applied statistics with Hebei GEO University, China. He has actively participated in numerous scientific research projects and academic skill competitions about big data and mathematical modeling, including two national-level innovation and entrepreneurship projects, one provincial-level project, and several university-level projects, earning 22 national-level awards, 12 provincial-level awards, one city/district-level award, and 14 university-level awards. He has also published over 16 research papers and holds nine authorized utility model patents in the fields of system science and big data technology. His main research interests include big data mining and analysis, machine learning, deep learning, data visualization, and time series analysis. His representative achievements include: 1) multimodal gesture recognition with spatiotemporal features fusion based on YOLOv5 and MediaPipe. *International Journal of Pattern Recognition and Artificial Intelligence*. (SCI & EI); and 2) study on the relationship between Hang Seng Index and FTSE CHI Index in Hong Kong: Evidence from VAR Model. Proceedings of the 2nd International Conference on Business and Policy Studies Series: *Advances in Economics, Management and Political Sciences*. (CPCI).



RUFEI ZHANG received the Ph.D. degree in statistics from the School of Statistics, Renmin University of China (RUC).

He is currently a Professor and a Master's Supervisor in statistics and applied statistics with Hebei GEO University, China, as well as a Visiting Professor at Wonkwang University, South Korea. His main research interests include: 1) dynamic effects of oil price shocks on exchange rates—from a time-varying perspective. *Sustainability*. (SSCI); 2) relaxed adaptive lasso and its asymptotic results. *Symmetry*. (SCI); 3) partisan conflict, national security policy uncertainty and tourism. *Sustainability*. (SSCI); 4) robust variable selection based on relaxed lad lasso. *Symmetry*. (SCI); and 5) analysis of the spatial differentiation and promotion potential for agricultural eco-efficiency—evidence of pollution's strong disposability. *International Journal of Environmental Research and Public Health*. (SSCI). He serves as a Council Member of the National Statistical Society of China, a Council Member of Hebei Provincial Statistical Society, the Deputy Secretary-General of Hebei Provincial Input-Output Association, the Deputy Secretary-General of the National Input-Output and Big Data Research Association, and a member of the International Input-Output Association.



WENXIN CAO received the master's degree in foreign languages and literature from Southwest Jiaotong University (SWJTU), China.

She has participated in two national-level projects and one university-level project, and published two monographs and two papers in the fields of applied linguistics and deep learning. Her main research interests include translation, computational linguistics, corpus linguistics, sociolinguistics, and psycholinguistics. The representative achievements include: 1) unpacking changing emotions in multiple contexts: idiodynamic study of college students' academic emotions. *International Review of Applied Linguistics in Language Teaching*. (SSCI); and 2) multimodal gesture recognition with spatio-temporal features fusion based on YOLOv5 and MediaPipe. *International Journal of Pattern Recognition and Artificial Intelligence*. (SCI & EI).