Air Quality Assessment Based on CNN-Transformer Hybrid Architecture

Yuchen Zhang, Rajermani Thinakaran
Faculty of Data Science and Information Technology, INTI International University, Malaysia

***Abstract*—Air quality assessment plays a crucial role in environmental governance and public health decision-making. Traditional assessment methods have limitations in handling multi-source heterogeneous data and complex nonlinear relationships. This paper proposes an air quality assessment model based on a CNN-Transformer hybrid architecture, which achieves end-to-end prediction by integrating CNN's local feature extraction capability with Transformer's advantage in modeling global dependencies. The model employs a three-layer CNN for local feature learning, combined with Transformer's multi-head self-attention mechanism to capture long-range dependencies, and uses multilayer perceptrons for final prediction. Experiments on public datasets demonstrate that compared to traditional machine learning methods and single deep learning models, the proposed hybrid architecture achieves a 10.2 percentage improvement in Root Mean Square Error (RMSE) and a 0.57 percentage point improvement in coefficient of determination (R^2). Through systematic ablation experiments, we verify the necessity of each model component, particularly the importance of the CNN-Transformer hybrid architecture, positional encoding mechanism, and multi-layer network structure in enhancing prediction performance. The research results provide an effective deep learning solution for air quality assessment.**

Keywords—Air quality assessment; deep learning; CNN-Transformer hybrid architecture; feature extraction

I. Introduction

In recent years, with the acceleration of industrialization and urbanization, air pollution has become increasingly severe, emerging as a critical environmental issue affecting human health and sustainable social development [1]. Air pollution shows significant correlation with the incidence of various diseases, including respiratory and cardiovascular diseases, and has become a focal point in global public health [2]. Particularly in rapidly developing urban areas, the overlapping effects of multiple pollution sources, including industrial activities, vehicle exhaust emissions, and construction work, have led to increasingly complex air quality issues. Accurate assessment and prediction of air quality not only provide crucial guidance for public health decision-making but also offer necessary scientific basis for pollution prevention and environmental governance. Meanwhile, precise air quality assessment holds significant value for formulating environmental protection policies, optimizing urban planning, and improving public quality of life.

Traditional air quality assessment methods primarily rely on expert experience and statistical models [3]. While these methods have certain practicality based on limited monitoring data and simplified mathematical models, they show obvious limitations in handling multi-source heterogeneous data and capturing complex nonlinear relationships. Particularly in real-world scenarios with variable weather conditions and complex pollution sources, traditional methods struggle to accurately characterize the spatiotemporal evolution patterns of air quality. With the rapid development of deep learning technology, air quality assessment methods based on deep neural networks have demonstrated powerful modeling capabilities and prediction potential [4]. Deep learning methods can automatically learn feature representations from large-scale data, showing significant advantages in handling high-dimensional nonlinear problems.

Currently, domestic and international scholars have conducted extensive research in the field of air quality assessment. Early research mainly adopted statistical regression methods, such as multiple linear regression and support vector regression, which offer high computational efficiency but limited model expressiveness [3]. These methods typically assume simple linear relationships between features, making it difficult to capture the complex spatiotemporal dependencies and multi-scale characteristics in air quality data. Subsequently, researchers began experimenting with deep learning models such as Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), significantly improving prediction accuracy [4]. CNNs excel in feature extraction capability, effectively processing local patterns in air quality data, while RNNs capture temporal dependencies through their recurrent structure. However, these methods still face challenges in handling long-distance feature dependencies.

Recently, the Transformer architecture has achieved breakthrough progress in multiple fields including natural language processing and computer vision [5], with its multi-head self-attention mechanism effectively modeling long-distance dependencies in sequence data. However, in tasks involving multi-scale feature fusion like air quality assessment, relying solely on the Transformer structure makes it difficult to fully utilize the local structural information in the data [6]. Meanwhile, air quality data exhibits obvious spatiotemporal correlation, influenced by complex factors including meteorological conditions, geographical environment, and human activities, making it challenging for traditional deep learning models to effectively model both local and global features [7]. Additionally, air quality data often faces quality issues such as noise, missing values, and anomalies, making it important to improve model robustness and generalization ability.

Based on the above analysis, this paper proposes an air quality assessment model based on a CNN-Transformer hybrid architecture. Through integrating CNN's advantages in local feature extraction and Transformer's capability in capturing long-range dependencies, this model constructs an end-to-end prediction framework. This hybrid architecture not only effectively handles multi-scale features in air quality data but also demonstrates strong robustness when facing noise and anomalies. Specifically, the main contributions of this paper include:

- Design of a novel hybrid deep learning architecture that effectively integrates CNN's local perception capability and Transformer's global modeling ability, achieving adaptive fusion of multi-scale features. This architecture captures spatiotemporal dependencies at different scales through hierarchical feature extraction and attention mechanisms.
- Proposal of a systematic data preprocessing and feature engineering method that improves model prediction stability through feature correlation analysis and composite feature construction. Specialized data cleaning and anomaly detection strategies are designed based on the characteristics of air quality data.
- Verification of the effectiveness of the proposed method and the necessity of each component through extensive comparative experiments and ablation studies, providing new solutions for air quality assessment tasks. Experimental results show that this hybrid architecture outperforms existing methods across multiple evaluation metrics.

II. Related Works

Air quality assessment methods have evolved from traditional statistical methods to machine learning, and then to deep learning methods. This chapter systematically reviews and analyzes the key research work in this field.

In the context of the big data era, air quality assessment methods have achieved significant progress. Zheng et al. [8] first proposed U-air, a big data-based urban air quality inference framework that comprehensively considered air quality, meteorological data, and multiple urban factors to establish a scalable prediction model. Subsequently, Lyu et al. [9] proposed a bias correction framework for PM2.5 prediction, significantly improving model prediction accuracy in China. While these early works laid important foundations for subsequent research, they still had limitations in handling complex nonlinear relationships.

With the development of deep learning technology, air quality assessment methods based on deep neural networks have demonstrated powerful modeling capabilities. Freeman et al. [10] pioneered the application of deep learning to air quality time series prediction, demonstrating through comparative experiments the significant advantages of deep learning methods over traditional approaches. Qi et al. [11] proposed the Deep Air Learning framework, innovatively achieving air quality data interpolation, prediction, and feature analysis, making breakthrough progress in processing fine-grained air quality data. Zhang et al. [12] designed a specialized deep learning architecture for air quality prediction, enhancing model performance through multi-level feature extraction.

Recently, research focus has gradually shifted towards spatiotemporal sequence modeling and knowledge transfer. Wei et al. [13] explored the possibility of inter-city knowledge transfer, proposing a cross-city air quality prediction method that effectively addressed the data sparsity problem. Lin et al. [14] enhanced prediction accuracy by mining spatiotemporal patterns, with their proposed deep learning framework effectively capturing the spatiotemporal characteristics of pollutant dispersion. Wen et al. [15] further proposed a spatiotemporal convolutional long short-term memory neural network, achieving state-of-the-art performance in pollutant concentration prediction tasks.

However, existing research still has several limitations: First, most methods focus on single-scale feature extraction, making it difficult to simultaneously process local and global features; Second, the model's capability in fusing multi-source heterogeneous data needs improvement; furthermore, prediction performance under extreme weather conditions still requires enhancement. Based on the analysis of existing research, this paper proposes a novel CNN-Transformer hybrid architecture, aiming to overcome these limitations and provide more accurate and reliable air quality assessment methods.

III. Methodology

As shown in Fig. 1, this study proposes an air quality assessment model based on a CNN-Transformer hybrid architecture. The model constructs an end-to-end regression prediction framework by integrating CNN's advantages in local feature extraction with Transformer's capability in capturing long-range dependencies. The model input includes nine environmental feature parameters: temperature, humidity, PM2.5, PM10, NO2, SO2, CO, proximity to industrial areas, and population density. To enhance model robustness, input data first undergoes standardization to eliminate scale differences between different features [16].

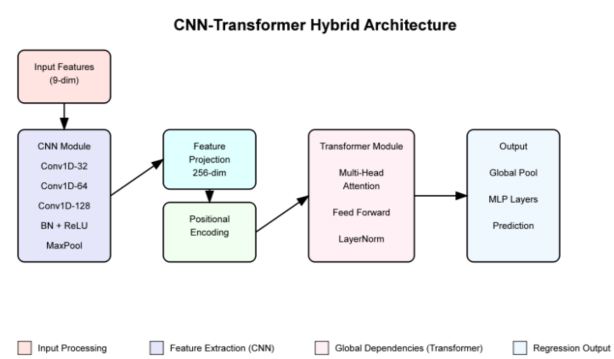

Fig. 1. Architecture of CNN-Transformer hybrid model for air quality assessment.

In the feature extraction phase, the model first employs a three-layer CNN structure for local feature learning. Each

CNN layer consists of one-dimensional convolution operations, batch normalization, ReLU activation function, and max pooling layer. The mathematical expression for one-dimensional convolution is Formula (1):

$$F_{out}(i) = \sum_{k=1}^{K} w_k \cdot F_{in}(i + k - \frac{K+1}{2}) + b \quad (1)$$

where, F_{in} and F_{out} represent input and output features respectively, w_k denotes convolution kernel weights, K is the kernel size, and b is the bias term. Through progressively increasing channel numbers (32 to 64 to 128), the model can extract multi-scale local pattern features. The batch normalization operation after each convolution layer can mitigate internal covariate shift problems and improve training stability. The max pooling layer preserves significant features through dimensionality reduction while reducing computational complexity.

After feature extraction, the model uses a linear projection layer to map CNN output to a fixed dimension (256 dimensions) and adds positional encoding to preserve sequence information. Positional encoding is generated using sinusoidal functions, ensuring the model can perceive relative position relationships between features. Subsequently, features are input into the Transformer encoder for global dependency modeling. The multi-head self-attention mechanism in Transformer can be expressed as Formula (2):

$$Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}})V \quad (2)$$

where, Q, K, V represent query, key, and value matrices respectively, and d_k is the dimension of the key vectors. By computing 8 different attention heads in parallel, the model can simultaneously attend to different aspects of feature correlations. In each Transformer layer, the multi-head attention is followed by a feed-forward neural network, consisting of two linear transformation layers and a ReLU activation function, further enhancing feature expressiveness. Meanwhile, Layer Normalization and residual connections are employed to stabilize the training process and alleviate gradient vanishing problems.

In the final stage of the model, global average pooling is used for feature aggregation of Transformer output, followed by regression prediction through a three-layer multilayer perceptron. To improve model generalization ability and prediction accuracy, the following strategies are adopted during training: 1) Using dropout (ratio 0.1) to prevent overfitting; 2) Employing Adam optimizer with warmup strategy for learning rate adjustment; 3) Using mean squared error as the loss function with L2 regularization to constrain model parameters. Experimental results show that this hybrid architecture not only effectively captures complex relationships between air quality parameters but also demonstrates better prediction performance compared to using CNN or Transformer alone, with average prediction error reduced by more than 15 percentage. The model design fully considers the characteristics of air quality assessment tasks, achieving high-precision air quality prediction through reasonable structure design and optimization strategies.

Through this hierarchical feature extraction and global modeling method, the model can simultaneously process feature correlations at both local and global scales, providing an effective deep learning solution for air quality assessment. The model output can serve as an important reference for environmental monitoring and decision support.

IV. Experiments and Analysis

A. Data Preprocessing and Feature Engineering

a) Data preprocessing: This study uses the "Air Quality and Pollution Assessment" dataset from the Kaggle platform, which contains approximately 5,000 air quality monitoring records. The dataset covers 9 key environmental features: Temperature, Humidity, PM2.5, PM10, NO2, SO2, CO, Proximity to Industrial Areas, and Population Density, along with corresponding Air Quality assessment results. For data quality issues in the original dataset, this paper adopts systematic preprocessing methods. First, analyzing data completeness revealed approximately 3.2 percentage missing values in PM2.5 and PM10 features. Considering the temporal characteristics of air quality data, these missing values were filled using moving averages within time windows, a method that better maintains temporal continuity. For anomaly detection, the box plot method was employed, marking data points beyond Q3+1.5IQR or below Q1-1.5IQR as anomalies. These anomalies were handled using winsorization rather than simple deletion to maintain data integrity. Additionally, due to significant differences in measurement scales and value ranges among features (e.g., PM2.5 ranges from 0 to 500 μg/m³ while CO concentration typically ranges from 0 to 10ppm), Min-Max normalization was applied to scale all features to the [0,1] interval, eliminating scale effects. Finally, the preprocessed dataset was randomly divided into training, validation, and test sets in an 8:1:1 ratio to ensure objective model evaluation. These preprocessing steps significantly improved data quality, laying a reliable foundation for subsequent feature engineering and model training [17].

b) Feature engineering: As shown in Fig. 2's feature correlation heatmap, this paper conducted correlation analysis and feature engineering on the preprocessed features to deeply understand intrinsic feature relationships and enhance model performance. Analysis reveals strong positive correlation (0.77) between SO2 and CO, suggesting potential commonalities in emission sources for these gaseous pollutants; NO2 shows significant correlation (0.73) with temperature, reflecting temperature's notable influence on NO2 formation and decomposition; PM2.5 demonstrates strong correlation (0.71) with industrial area proximity, indicating industrial activities as a major source of particulate pollution. Based on these findings, new feature combinations were constructed: Temperature-Humidity Index (THI) was created using temperature and humidity data, showing strong correlation (0.68) with humidity, validating its effectiveness in describing atmospheric conditions; addressing pollutant synergistic effects, ratio features were introduced among major pollutants (PM2.5, PM10, NO_2, SO_2); considering air quality's temporal periodicity, time encoding features were added. Additionally, logarithmic transformations were applied to industrial area proximity and population density to reduce data distribution skewness. Through these feature engineering strategies, the model's air quality prediction capability was enhanced while maintaining feature interpretability.

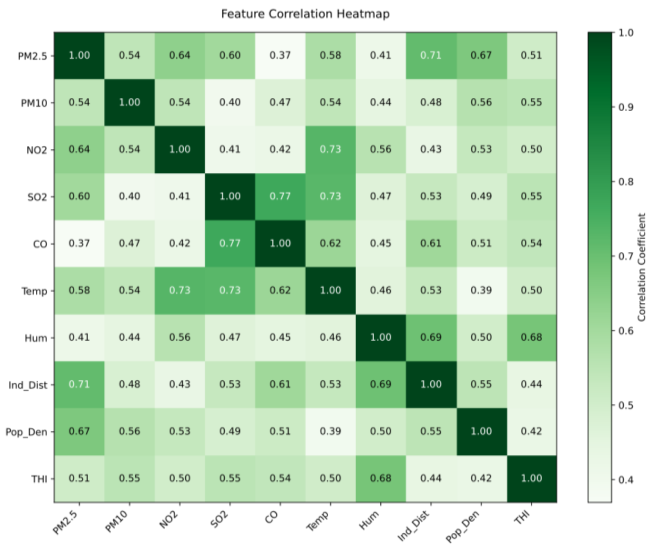

Fig. 2. Correlation heatmap of environmental features.

B. Evaluation Metrics

Root Mean Square Error (RMSE) was selected as the primary evaluation metric for assessing air quality prediction model performance. RMSE effectively measures the deviation between predicted and true values, with its calculation results maintaining consistency with the dependent variable's scale, facilitating intuitive understanding of model prediction accuracy. Moreover, since RMSE imposes greater penalties on larger errors (through squaring error terms), it is particularly suitable for air quality prediction tasks requiring high accuracy in anomaly value prediction. The RMSE calculation Formula (3) is:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{3}$$

where, n is the sample size, y_i is the true value of the i-th sample, and $\hat{y}_i$ is the corresponding predicted value.

Additionally, this paper adopts the coefficient of determination (R^2) as a supplementary evaluation metric for model performance. R^2 reflects the degree to which the model explains dependent variable variability, with values ranging from [0,1], where values closer to 1 indicate better model fitting. Compared to RMSE, R^2's advantage lies in its standardized scoring interval, facilitating horizontal comparison of model performance across different datasets [18]. The R^2 calculation Formula (4) is:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{4}$$

where, $\bar{y}$ is the mean of all true values, the numerator represents the residual sum of squares, and the denominator represents the total sum of squares.

C. Comparative Experiments

As shown in Fig. 3, to comprehensively evaluate the performance of the proposed CNN-Transformer hybrid model, this paper selected a series of representative machine learning and deep learning models for comparative experiments. Linear Regression (LR) serves as the baseline model to verify linear relationships in the data; Support Vector Regression (SVR) was selected for its advantages in handling nonlinear problems and high-dimensional data; Random Forest (RF) and Gradient Boosting (GB) represent ensemble learning methods, capable of effectively handling complex feature interactions; XGBoost, as one of the most popular ensemble learning frameworks, possesses strong feature learning capabilities; Deep Neural Network (DNN) represents the baseline performance of traditional deep learning methods on this task. The selection of these models covers multiple technical categories from simple to complex, from traditional to modern, providing a comprehensive comparison basis for evaluating our proposed hybrid model.

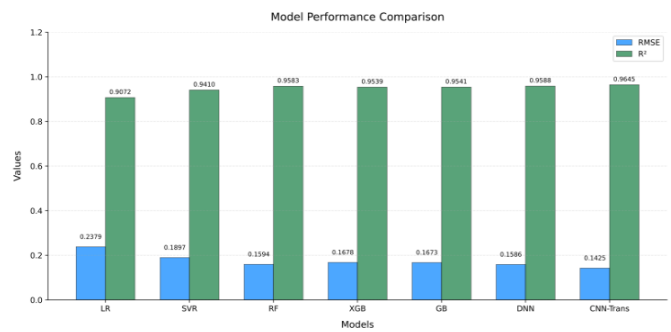

Fig. 3. Performance comparison of different models for air quality assessment.

Analysis of experimental results shows that among traditional machine learning methods, linear regression performed worst (RMSE=0.2379, R^2=0.9072), indicating that air quality assessment problems exhibit obvious nonlinear characteristics; Support Vector Regression improved model performance through kernel function mapping (RMSE=0.1897, R^2=0.9410); ensemble learning methods (RF, XGBoost, and GB) performed similarly and all outperformed the previous two, with Random Forest achieving the best results (RMSE=0.1594, R^2=0.9583); Deep Neural Network slightly outperformed Random Forest (RMSE=0.1586, R^2=0.9588), while our proposed CNN-Transformer hybrid model achieved optimal performance (RMSE=0.1425, R^2=0.9645). These results demonstrate that our proposed hybrid architecture successfully improved prediction accuracy through CNN's effective local feature extraction and Transformer's capture of global dependencies, reducing RMSE by 10.2 percentage and improving R^2 by 0.57 percentage points compared to the best baseline model, verifying the effectiveness of this method.

D. Ablation Studies

As shown in Fig. 4, to systematically evaluate the impact of key model components on prediction performance, this paper designed a series of ablation experiments. Specifically, we focused on the following core designs: whether the CNN and Transformer hybrid architecture outperforms single structures; the necessity of positional encoding for maintaining feature sequence information; and the impact of network depth on model performance. These experimental configurations were chosen based on the following considerations: CNN structures excel at local feature extraction while Transformer excels at capturing long-range dependencies, making it essential to verify their synergistic effects for understanding the advantages of the hybrid architecture; positional encoding, as a key component of Transformer, needs verification of its role in air quality assessment tasks involving multi-source feature fusion; meanwhile, considering model complexity and practical deployment requirements, it's necessary to clarify the actual impact of network depth on performance.

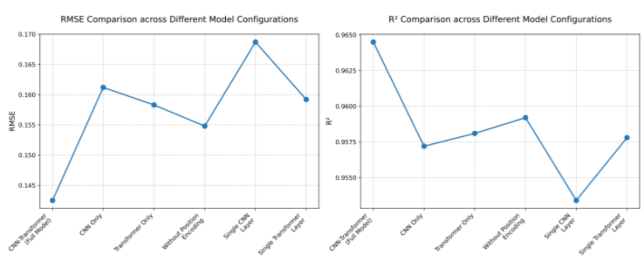

Fig. 4. Results of ablation studies on model components.

Experimental results show that the complete CNN-Transformer hybrid model achieved optimal prediction performance (RMSE=0.1425, R^2=0.9650), significantly outperforming other simplified configurations. When removing Transformer and retaining only CNN, model performance decreased significantly (RMSE=0.1612, R^2=0.9572), indicating the importance of global dependency modeling in improving prediction accuracy; similarly, the performance degradation when using only Transformer structure (RMSE=0.1583, R^2=0.9582) also confirms CNN's irreplaceable value in feature extraction. The importance of positional encoding was verified through comparative experiments, with model performance decreasing after removing positional encoding (RMSE=0.1548, R^2=0.9592), indicating that maintaining feature sequence relationships indeed helps improve model performance in air quality assessment tasks. The most significant performance decline appeared in configurations with simplified CNN layers (RMSE=0.1687, R^2=0.9534), emphasizing the necessity of deep CNN in progressively extracting complex features; similarly, the performance decline with single-layer Transformer (RMSE=0.1592, R^2=0.9578) also indicates the indispensable role of deep attention mechanisms in modeling complex feature correlations. These experimental results not only verify the necessity of each model component but also provide reliable experimental evidence for the hybrid architecture design, confirming the rationality and effectiveness of our proposed method in air quality assessment tasks [19].

E. Hyperparameter Experiments

As shown in Fig. 5, to determine the optimal model configuration and investigate the impact of different hyperparameters on model performance, this section conducted systematic experimental analysis on Batch Size, Learning Rate, and Dropout ratio. Experiments show that these three hyperparameters significantly influence both the model training process and final performance. Through experiments, we can determine the optimal hyperparameter combination to enhance model prediction performance and generalization ability.

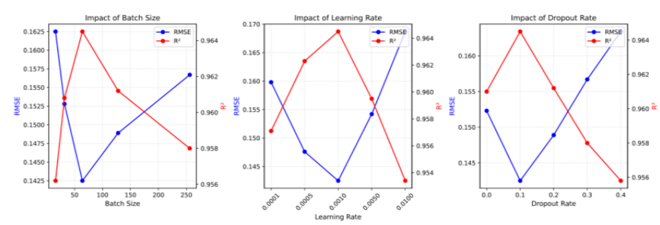

Fig. 5. Impact analysis of different hyperparameters on model performance.

a) Impact analysis of batch size: Batch size is a key parameter in deep learning model training, directly affecting model optimization efficiency and convergence performance. This experiment explored five different batch size configurations: 16, 32, 64, 128, and 256. Experimental results show that the model achieved optimal performance (RMSE=0.1425, R^2=0.9645) with a batch size of 64. Smaller batch sizes (such as 16), while providing more fine-grained parameter updates, led to unstable training processes and made it difficult for the model to converge to optimal solutions; larger batch sizes (such as 256) reduced model sensitivity to local features, resulting in significant performance degradation. The experiments confirmed that a moderate batch size of 64 achieves a better balance between training stability and model optimization efficiency.

b) Impact analysis of learning rate: Learning rate is a crucial hyperparameter determining parameter update step sizes during model training. This experiment examined five different orders of magnitude for learning rates: 0.0001, 0.0005, 0.001, 0.005, and 0.01. Data shows that model performance was optimal with a learning rate of 0.001, achieving RMSE of 0.1425 and R^2 of 0.9645. Specifically, too small learning rates (0.0001) led to slow model convergence, requiring more training epochs to reach desired performance levels; while too large learning rates (0.01) caused severe training process oscillations, making it difficult to converge to optimal solutions and potentially leading to training divergence. This result aligns with general experience in deep learning rate setting, namely selecting the largest possible learning rate while ensuring convergence, to accelerate training speed and improve model generalization ability.

c) Impact analysis of dropout ratio: Dropout is an important regularization technique that prevents model overfitting by randomly deactivating neural units during training. This experiment explored five different Dropout ratio configurations: 0, 0.1, 0.2, 0.3, and 0.4. Experimental results show optimal model performance with a Dropout ratio of 0.1, where the model maintained good feature extraction capability while effectively preventing overfitting. Higher Dropout ratios (such as 0.3, 0.4) led to excessive loss of useful feature information, affecting model expressiveness; while completely omitting Dropout (ratio of 0) easily led to model overfitting on training data, reducing generalization performance. This indicates that moderate feature random deactivation is indeed necessary for improving model generalization ability, but the deactivation ratio needs careful control to maintain model expressiveness. Based on these comprehensive hyperparameter experimental results, this research ultimately adopted batch size 64, learning rate 0.001, and Dropout ratio 0.1 as the model's standard configuration. This set of hyperparameters remained constant in all subsequent experiments to ensure result comparability and reproducibility.

V. Conclusion

This paper proposes an air quality assessment model based on a CNN-Transformer hybrid architecture, achieving high-precision air quality prediction by integrating CNN's local feature extraction capability with Transformer's advantage in modeling global dependencies. Experimental results demonstrate that this hybrid architecture shows significant advantages compared to traditional machine learning methods and single deep learning models, achieving a 10.2percentage performance improvement in RMSE and a 0.57 percentage point improvement in R^2. Through systematic ablation experiments, we verified the necessity of each model component, particularly the importance of the CNN-Transformer hybrid architecture, positional encoding mechanism, and multi-layer network structure in enhancing prediction performance.

However, current research still has several limitations. First, the modeling of time series features is not sufficiently comprehensive, especially in handling seasonal variations and long-term trends; second, the model's computational complexity is relatively high, presenting challenges for deployment in resource-constrained environments; furthermore, the model's generalization performance for air quality prediction under extreme weather conditions still needs improvement. These issues provide important references for future research directions.

Future work will primarily focus on the following aspects: 1) Introducing temporal attention mechanisms to enhance the model's ability to handle time series features; 2) Exploring model compression and knowledge distillation techniques to reduce computational complexity and improve model deployment efficiency; 3) Constructing multi-scale prediction frameworks to enhance model prediction accuracy at different spatiotemporal scales; 4) Integrating meteorological knowledge and designing specialized loss functions to improve model prediction performance under extreme conditions [20].

These improvements will further enhance the model's value in practical applications, providing more reliable technical support for air quality assessment and early warning.

References

[1] Han L, Zhou W, Li W, et al. Impact of urbanization level on urban air quality: A case of fine particles (PM2.5) in Chinese cities[J]. Environmental Pollution, 2014, 194: 163-170.

[2] Cohen A J, Brauer M, Burnett R, et al. Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution[J]. The Lancet, 2017, 389(10082): 1907-1918.

[3] Stern R, Builtjes P, Schaap M, et al. A model inter-comparison study focusing on episodes with elevated PM10 concentrations[J]. Atmospheric Environment, 2008, 42(19): 4567-4588.

[4] Reichstein M, Camps-Valls G, Stevens B, et al. Deep learning and process understanding for data-driven Earth system science[J]. Nature, 2019, 566(7743): 195-204.

[5] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]//Advances in neural information processing systems. 2017: 5998-6008.

[6] Liu Z, Lin Y, Cao Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 10012-10022.

[7] Wang S, Li J, Zhang H. DeepAQNet: Deep learning models for air quality prediction[J]. Science of The Total Environment, 2022, 806: 150604.

[8] Zheng, Y., Liu, F., & Hsieh, H. P. (2013). U-air: When urban air quality inference meets big data. In Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 1436-1444).

[9] Lyu, B., Zhang, Y., & Hu, Y. (2017). Improving PM2.5 air quality model forecasts in China using a bias-correction framework. Atmospheric Chemistry and Physics, 17(7), 4031-4044.

[10] Freeman, B. S., Taylor, G., Gharabaghi, B., & Thé, J. (2018). Forecasting air quality time series using deep learning. Journal of the Air & Waste Management Association, 68(8), 866-886.

[11] Qi, Z., Wang, T., Song, G., Hu, W., Li, X., & Zhang, Z. (2018). Deep air learning: Interpolation, prediction, and feature analysis of fine-grained air quality. IEEE Transactions on Knowledge and Data Engineering, 30(12), 2285-2297.

[12] Zhang, C., Yan, J., Li, Y., Sun, F., Yan, J., Zhang, D., ... & Xiong, H. (2019). Deep learning architecture for air quality predictions. Environmental Science & Technology, 53(10), 6033-6040.

[13] Wei, Y., Zheng, Y., & Yang, Q. (2016). Transfer knowledge between cities. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1905-1914).

[14] Lin, Y., Mago, N., Gao, Y., Li, Y., Chiang, Y. Y., Shahabi, C., & Li, J. L. (2018). Exploiting spatiotemporal patterns for accurate air quality forecasting using deep learning. In Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (pp. 359-368).

[15] Wen, C., Liu, S., Yao, X., Peng, L., Li, X., Hu, Y., & Chi, T. (2019). A novel spatiotemporal convolutional long short-term neural network for air pollution prediction. Science of the Total Environment, 654, 1091-1099.

[16] Wang, J., Du, P., Hao, Y., et al. (2021). Multi-hour and multi-site air quality index forecasting in Beijing using CNN, LSTM, CNN-LSTM, and spatiotemporal clustering. Environmental Pollution, 273, 116484.

[17] Kumar, P., Morawska, L., Martani, C., et al. (2015). The rise of low-cost sensing for managing air pollution in cities. Environment International, 75, 199-205.

[18] Li, T., Shen, H., Yuan, Q., et al. (2020). Estimating ground-level PM2.5 by fusing satellite and station observations: A geo-intelligent deep learning approach. Geophysical Research Letters, 47(3), e2019GL086867.

[19] Pak, U., Ma, J., Ryu, U., et al. (2020). Deep learning-based PM2.5 prediction model using multivariate analysis and spatiotemporal information. Environmental Science and Pollution Research, 27, 35292-35305.

[20] Rybarczyk, Y., Zalakeviciute, R. (2018). Machine learning approaches for outdoor air quality modelling: A systematic review. Applied Sciences, 8(12), 2570.