High-Precision Urban Air Quality Prediction Using a LSTM-Transformer Hybrid Architecture

Yiming Liu[1], Mcxin Tee[2], Liangyan Lu[3], Fei Zhou[4], Binggui Lu[5]
Faculty of Business and Communications, INTI International University, Malaysia[1, 2]
Accounting and Finance Department, Yunnan College of Business Management, Kunming 655000, China[3]
President's Office, Shinawatra University, Pathum Thani 12160, Thailand[4]
Faculty of Education, Shinawatra University, Pathum Thani 12160, Thailand[5]

Abstract*—With the acceleration of urbanization, accurate air quality prediction is crucial for environmental governance and public health risk management. Existing prediction methods still face challenges in handling complex time-series dependencies and multi-scale features. In this paper, a hybrid deep learning architecture (LT-Hybrid) based on LSTM and Transformer is proposed for high-precision air quality prediction. The model captures the long-term dependencies of time-series data through a two-layer LSTM structure, models the complex interactions among different environmental factors using a multi-head self-attention mechanism, and improves the training stability through a combination of residual connections and layer normalization. Experiments on an urban air quality dataset, containing nine dimensions of environmental characteristics such as temperature, humidity, PM2.5, etc., show that the LT-Hybrid model achieves an RMSE of 0.1021 and an R² of 0.9382, reducing prediction errors by 13.0% and 5.1% compared to benchmark models of traditional LSTM and XGBoost, respectively. Accurate prediction of air quality indicators provides timely risk assessment for respiratory diseases and cardiovascular conditions, enabling proactive public health interventions. Through systematic ablation experiments and hyperparameter analysis, the validity of each core component of the model is verified, providing a high-precision prediction scheme for environmental monitoring and health risk assessment.

***Keywords*—Air quality; deep learning; LSTM; transformer; multi-head attention mechanism; temporal prediction; health risk**

I. Introduction

Air quality has become a key issue in modern urban development and public health management. With the acceleration of industrialization and urbanization, the spatial and temporal distribution of air pollutants is becoming more and more complex, and the interaction mechanisms between pollutants are more difficult to capture. Accurate air quality prediction not only provides scientific decision support for environmental regulators, but also helps the public to take protective measures in time, which is of great practical significance for improving public health and quality of life [1].

Traditional air quality prediction methods mainly include statistical modeling and numerical simulation. Statistical models such as autoregressive integral sliding average (ARIMA) are computationally efficient and easy to implement, but it is difficult to characterize the nonlinear relationship and long-term dependence between pollutants [2]; numerical simulation models such as community multiscale air quality model (CMAQ) take into account the detailed atmospheric physicochemical processes, but it is computationally expensive and requires a large number of accurate input parameters [3]. In recent years, with the booming development of deep learning techniques, neural network-based prediction methods have shown significant advantages. Among them, Long Short-Term Memory (LSTM) networks are widely used in time-series prediction tasks due to their unique gating mechanism that can effectively capture long-term dependencies, while Transformer models show excellent modeling capabilities when dealing with multivariate sequential data by virtue of their powerful self-attention mechanism [4].

However, existing deep learning methods still face three main challenges in the air quality prediction task: first, although a single LSTM model can model the temporal dependence, it is difficult to effectively capture the complex interactions between different environmental factors; second, the computational complexity of the standard Transformer increases significantly with the length of the sequences when dealing with long sequential data, which restricts its use in high-frequency environmental monitoring data analysis; finally, the common non-stationarity and multi-scale characteristics of environmental data also bring a severe test to the generalization ability of the prediction model [5].

To address the above problems, this paper proposes a hybrid architecture (LT-Hybrid) based on LSTM and Transformer for air quality prediction. The main contributions of this study include 1) proposing a novel hybrid deep learning architecture, which significantly improves the prediction performance by fusing the sequence modeling capability of LSTM and the feature interaction capability of the multi-head self-attention mechanism, reducing the prediction error by 13.0% and 5.1% compared to the benchmark models such as the traditional LSTM and XGBoost, respectively, and 2) designing a two-layer LSTM with a four-headed cascade structure of the attention mechanism, which realizes the adaptive extraction of multi-scale features and enables the model to reach 0.9382 in the R² evaluation index, which improves 1.66 percentage points compared to a single model; 3) by introducing the combined design of residual linkage and layer normalization, which effectively solves the training problem of the deep network, the ablation experiments show that this design reduces the RMSE of the model from 0.1142 to 0.1021, which improves the prediction stability. The experimental results show that the proposed LT-Hybrid model can effectively deal with the complex temporal dependencies in air quality prediction, and provides a high-precision prediction scheme for the field of environmental monitoring.

II. Related Work

A. Traditional Air Quality Prediction Methods

Air quality prediction studies first used statistical methods. Autoregressive integrated sliding average model (ARIMA) became the main tool for early air quality prediction due to its good performance in time series analysis. Qi et al [6] applied an improved ARIMA model to predict PM2.5 concentration in Beijing, and enhanced the model performance by introducing seasonal adjustment. Another important class of methods is prediction models based on numerical simulation, such as the community multi-scale air quality model (CMAQ). Zhang et al [7] applied the coupled WRF-CMAQ model to regional-scale air quality prediction, which is able to take into account the detailed atmospheric physicochemical processes but is computationally expensive and has stringent requirements on the quality of input data. In addition, machine learning methods such as Support Vector Regression (SVR) and Random Forest (RF) have been widely applied to air quality prediction. Zhai et al [8] constructed a multi-objective prediction framework based on XGBoost, which demonstrated the advantages of dealing with nonlinear relationships.

B. Deep Learning Based Prediction Methods

In recent years, deep learning has made significant progress in the field of air quality prediction. Recurrent neural network (RNN) and its variant LSTM have become a research hotspot due to its ability to effectively process sequential data. Wen et al [9] proposed a prediction model based on bidirectional LSTM, which improves the prediction accuracy by simultaneously considering the information of historical and future time steps. With the development of deep learning technology, Tao et al [10] proposed a deep learning model based on a one-dimensional convolutional network and a bidirectional GRU, which improves the prediction accuracy by effectively extracting spatio-temporal features. In addition, one-dimensional convolutional neural network (1D-CNN) has been demonstrated to have unique advantages in processing environmental time-series data. Huang et al [11] applied deep residual network to air quality prediction, which effectively mitigated the gradient vanishing problem through jump connections.

C. Hybrid Modeling and Multi-Source Data Fusion

In order to fully utilize the advantages of different models, researchers have begun to explore hybrid modeling approaches. Yi et al [12] proposed a deep distributed fusion network that significantly improves the prediction performance by fusing heterogeneous urban data to capture all influential factors. In terms of feature extraction, Freeman et al [13] proposed a novel deep learning architecture that improves prediction accuracy through multi-level feature extraction and fusion. Another important research direction is to introduce the attention mechanism for feature selection. Liang et al [14] proposed a deep learning model based on spatio-temporal attention, which is able to adaptively learn the importance of different spatio-temporal features, providing a new idea for air quality prediction. In addition, Yu et al [15] explored an air quality prediction method based on graph neural networks, which achieves high-precision prediction on a regional scale by modeling the spatial correlation relationship between monitoring stations.

These related works have laid an important foundation for the LT-Hybrid model proposed in this paper. Although existing studies have made progress in different aspects, there are still challenges in dealing with complex temporal dependencies and multi-scale feature fusion, which are the directions of focus and improvement in this paper.

III. Methodology

A. Problem Statement

Accurate prediction for urban air quality is one of the key tasks in environmental monitoring and management. In this paper, air quality prediction is modeled as a time-series prediction problem: given environmental monitoring data from the past 24 time steps, including nine characteristic dimensions such as temperature, humidity, PM2.5, PM10, NO2, SO2, CO concentration, and the distance to the industrial area and population density, we predict the target air quality indicators for the next time step. This prediction task is obviously challenging: first, the environmental data exhibit complex time-dependence and potential interactions among different pollutants; second, the air quality is affected by a combination of factors, including both dynamic changes in meteorological conditions and cyclical patterns of human activities; and lastly, the environmental data often exhibit nonlinear and non-smooth characteristics, which puts higher demands on the prediction model's generalization ability puts forward higher requirements. Therefore, it is of great practical significance to design a prediction model that can effectively capture these complex patterns.

Formally, if the input feature at the t^{-th} time step is denoted as $x_t \in \mathbb{R}^9$, the prediction task can be formulated as follows: based on the observation sequence $\{x_{t-23}, x_{t-22}, \ldots, x_t\}$ predicts the target value y_{t+1}. where, the input features contain multi-dimensional information reflecting the current environmental conditions, and the prediction targets focus on specific air quality indicators. With this sliding window approach, the model can continuously predict future air quality and provide data support for environmental regulation and public health decision-making.

B. Model Architecture

The air quality prediction model proposed in this paper is a hybrid architecture based on LSTM and Transformer, which improves the prediction performance by combining the advantages of both models [16]. As shown in Fig. 1, the model mainly consists of an LSTM coding layer, a multi-head self-attention mechanism, a feed-forward neural network and a normalization layer. Each core component is described in detail below.

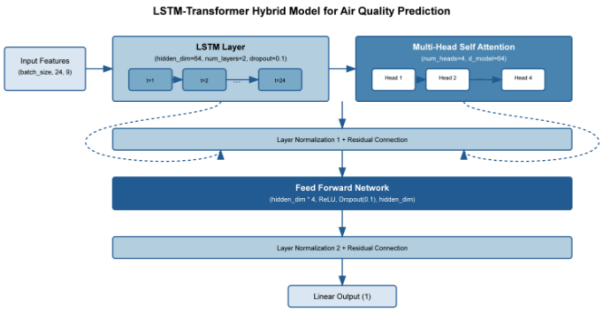
LSTM-Transformer Hybrid Model for Air Quality Prediction
Input Features
(batch_size, 24, 9)
LSTM Layer
(hidden_dim=64, num_layers=2, dropout=0.1)
t=1
t=2
...
t=24
Multi-Head Self Attention
(num_heads=4, d_model=64)
Head 1
Head 2
Head 4
Layer Normalization 1 + Residual Connection
Feed Forward Network
(hidden_dim * 4, ReLU, Dropout(0.1), hidden_dim)
Layer Normalization 2 + Residual Connection
Linear Output (1)

Fig. 1. Model architecture diagram.

C. LSTM Coding Layer

The LSTM coding layer is the first major component of the model for capturing long-term dependencies in temporal data [17]. Compared with traditional recurrent neural networks, LSTM can effectively mitigate the gradient vanishing problem and better maintain long-term memory through the gating mechanism. The layer adopts a two-layer LSTM structure (num_layers=2) with a hidden layer dimension of 128, and uses dropout=0.1 between layers to prevent overfitting. The core update process for each LSTM cell can be represented as:

$$c_t = f_t \odot c_{t-1} + i_t \odot tanh(W_c \cdot [h_{t-1}, x_t] + bc) \qquad (1)$$

In particular, the memory unit c_t realizes selective retention of historical information and selective reception of new information through the modulation of the forgetting gate f_t and the input gate i_t.

The LSTM layer receives a 9-dimensional sequence of input features (including environmental indicators such as temperature, humidity, PM2.5, etc.), and the length of the sequence is set to 24 time steps, which enables the model to make predictions based on data from the past 24 time units. This design fully takes into account the temporal characteristics of air quality data, as pollutant concentrations tend to exhibit obvious daily variation cycles and continuity. By cascading the two-layer LSTM, the model can capture the basic temporal patterns in the first layer and further extract the high-level temporal features in the second layer, so as to efficiently learn and memorize the important patterns in the environmental data at different time scales. Especially when dealing with environmental data with complex time dependencies, this cascaded feature extraction structure shows significant advantages.

D. Multi-Pronged Self-Attention Mechanisms

In order to enhance the model's ability to model the relationship between different time steps in temporal data, a multi-head self-attention mechanism was introduced after the LSTM layer [18]. Traditional attention mechanisms may assign too much attention weight to a single feature or time step, thus ignoring other potentially important information. The multi-head attention mechanism allows for simultaneous attention to different types of feature patterns by projecting the input into multiple subspaces [19]. The mechanism uses 4 attention heads (num_heads=4), each with a dimension of 16 (d_k=d_model/num_heads=64/4=16), and its core computational process can be represented as:

$$\mathrm{Attention}(Q, K, V) = \mathrm{softmax}(QK^T/\sqrt{d_k})V \qquad (2)$$

$$where, Q = HW^Q, K = HW^K, V = HW^V\mathrm{E}$$

The design of multiple heads of attention allows the model to learn feature associations in parallel in different representation subspaces. Each attention head can focus on capturing specific types of dependencies; for example, one head may focus on short-term correlations between temperature and humidity, while another may focus on long-term patterns of association between PM2.5 and other pollutants. Through this parallel processing mechanism, the model is able to model dependencies on multiple time scales simultaneously, capturing both localized patterns of rapid change as well as identifying global trends of long-term change, thus significantly improving the model's ability to understand and predict complex spatial and temporal patterns.

E. Residual Connections and Layer Normalization

A combination of two residual connections and layer normalization is used in the model, located after the multi-head attention layer and the feedforward network layer, respectively [15]. This design draws on the architectural features of Transformer and helps mitigate the problem of gradient vanishing in deep neural network training. Layer normalization helps to stabilize the training process, while residual connectivity maintains the information of the low-level features, allowing the model to better integrate different levels of feature representation. The use of this architecture significantly improves the training stability and convergence speed of the model.

F. Feedforward Neural Networks

After the multi-head attention layer, the model uses a feed-forward neural network for feature transformation. The network uses an expansion-contraction structure, where the feature dimensions are first expanded to four times their original size (hidden_dim*4), then nonlinearities are introduced via the ReLU activation function, and finally the dimensions are compressed back to their original size (hidden_dim). The network also uses dropout (rate of 0.1) to prevent overfitting. This component enables further abstraction and transformation of the features extracted by the attention mechanism, enhancing the expressive power of the model.

G. Output Layer

The final layer of the model is a linear output layer that maps the processed features to a single predicted value. This layer compresses the high-dimensional features extracted and transformed above into a one-dimensional output that directly predicts the target air quality indicator. With this end-to-end architectural design, the model is able to automatically learn the complex mapping relationships from the original input features to the final predicted values.

This hybrid architecture design takes full advantage of LSTM's strengths in sequence modeling and Transformer's strengths in feature interaction modeling, enabling the model to better handle complex time-series prediction tasks such as air quality prediction.

IV. Experiments

A. Data Preprocessing

In this study, we use the publicly available dataset "Urban Air Quality Dataset" from the Kaggle platform, which contains the environmental monitoring data of a city during the period of 2020-2023, totaling 5000 records. The dataset covers nine dimensions of environmental characteristics: temperature (℃), relative humidity (%), PM2.5 (μg/m³), PM10 (μg/m³), NO2 (μg/m³), SO2 (μg/m³), and CO (mg/m³), as well as two spatial characteristics: distance from the industrial area (km) and population density of the area (people/km²). The data sampling frequency was hourly, ensuring continuous monitoring of air quality changes.

As shown in Fig. 2, from the time-series distribution of the key features, the temperature values generally fluctuate between 20 and 40°C, reflecting obvious daily changes; the relative humidity has a large range of variation, fluctuating between 40 and 100%, and fluctuates more frequently; and the PM2.5 concentration shows a large fluctuation, with the baseline value between 0 and 40 μg/m³, but with obvious peaks (the highest reaching about 140 μg/m³), reflecting the fact that air quality can deteriorate significantly at certain points in time. The time-series change characteristics of these three key indicators indicate that the air quality of the city is affected by a combination of several environmental factors, showing a complex dynamic change pattern, which provides an important basis for the subsequent predictive modeling.

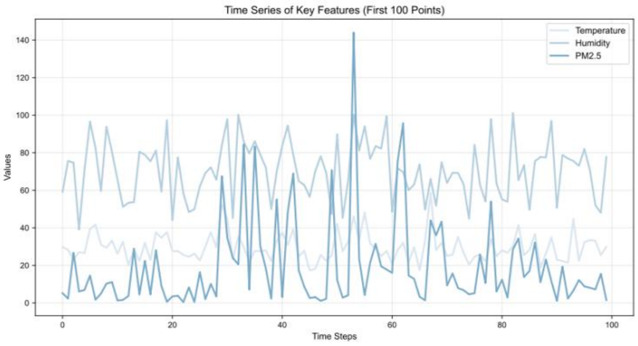

Fig. 2. Time series plot of key features.

B. Feature Engineering

In order to improve the training effect of the model, this paper carries out a series of pre-processing on the raw data. First, the data are processed for missing values, and the moving average method is used to fill in a small amount of missing monitoring data to ensure the continuity of the data. Second, the individual features are normalized using MinMaxScaler, which maps the data to the [0, 1] interval and eliminates the scale differences brought about by different units of measure. Finally, the sliding window method is used to construct the time series samples, and 24 hours are selected as the length of the input sequence, i.e., the data of the previous 24 hours are used to predict the air quality indicators of the next hour, so as to generate the input-output sample pairs required for model training.

C. Experimental Setup

For the experimental setup, we adopted a rigorous training-validation-testing framework. First, the processed dataset was randomly divided into a training set (4000 entries), a validation set (500 entries), and a testing set (500 entries) according to the ratio of 8:1:1, and ensured that the continuity of the time series was maintained during the division process. The model was trained using the Adam optimizer with the initial learning rate set to 0.001, and the learning rate was adjusted using the cosine annealing strategy. To prevent overfitting, a regularization strategy with dropout=0.1 was used during training, and an early stopping strategy was applied to the validation set, where, training was stopped when the validation loss did not improve within 10 consecutive epochs.

In terms of model hyperparameter configuration, the LSTM coding layer uses a two-layer structure (num_layers=2), and the hidden layer dimension is set to 128; the multi-head attention mechanism uses four attention heads (num_heads=4), each with a dimension of 16; the training batch size (batch_size) is set to 32, and the maximum number of training rounds (epochs) is 100. All experiments were conducted on workstations configured with NVIDIA RTX 3080 GPUs and implemented using the PyTorch 1.9.0 framework. To ensure the reliability of the experimental results, all experiments were repeated three times and the average value was taken as the final result.

D. Assessment of Indicators

In order to comprehensively evaluate the prediction performance of the model, this paper chooses the Root Mean Square Error (RMSE) as the main evaluation index, which can visually reflect the degree of deviation between the predicted values and the real values, and its calculation results are consistent with the scale of the dependent variable, which makes the evaluation results easier to understand and interpret. For regression problems such as air quality prediction, RMSE can clearly indicate the average level of prediction error, and its calculation formula is:

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

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

Meanwhile, this paper also adopts the coefficient of determination (R^2) as a supplementary assessment indicator. R^2 reflects the extent to which the model explains the variability of the dependent variable, and its value ranges from 0 to 1, with the closer it is to 1 indicating that the model's explanatory ability is stronger. The strength of this metric is that it can help us understand the model's ability to capture patterns of data variability, especially in assessing the model's grasp of long-term trends in air quality. The formula for R^2 is:

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

where, ȳ is the mean of the true values, a metric that effectively assesses the overall goodness of fit of the model by comparing the ratio of the model's prediction error to the variability of the data itself.

E. Comparative Experiments

As shown in Fig. 3, in order to comprehensively evaluate the performance of the proposed LT-Hybrid model, six representative machine learning and deep learning models are selected as benchmarks for comparative experiments in this paper [20]. These benchmark models include: support vector regression (SVR), which is a traditional machine learning method with good non-linear modeling capability; long-short-term memory network (LSTM), which is widely used in the field of time-series prediction; three integrated learning methods that have excellent performance in modeling environmental data, i.e., Random Forest (RF), XGBoost (XGB), and Gradient Boosting (GB); and as a deep learning representative of deep neural networks (DNNs).

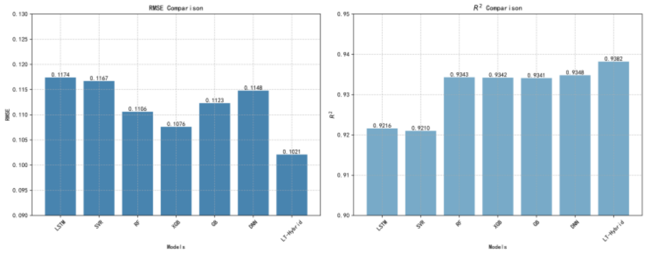

Fig. 3. Comparison experiment.

The experimental results show that the LT-Hybrid model achieves the optimal performance in both RMSE and R² evaluation metrics. In terms of the RMSE metric, the LT-Hybrid model achieves the lowest error of 0.1021, which reduces the prediction error by about 5.1% compared to the second best performing XGBoost model (RMSE of 0.1076), and improves the prediction error compared to the traditional LSTM (RMSE of 0.1174) and SVR (RMSE of 0.1167) by 13.0% and 12.4%. In terms of model fit goodness, the LT-Hybrid model has an R² value of 0.9382, indicating that the model is able to explain about 93.82% of the data variability, while the R² values of the other models are generally in the range of 0.92-0.93. Notably, the integrated learning methods (RF, XGB, and GB) outperformed the single model overall, reflecting the importance of model integration in modeling complex environmental data.

Through comparative experiments, it can be found that the advantages of the LT-Hybrid model mainly come from its unique hybrid architecture design. Compared with a single sequence modeling method (e.g., LSTM) or a traditional regression model (e.g., SVR), the hybrid architecture proposed in this paper effectively enhances the ability to capture complex relationships among environmental factors by integrating the long-term dependency modeling capability of LSTM and the feature interaction capability of Transformer, while maintaining the advantages of time-series modeling. This design not only improves the prediction accuracy of the model, but also enhances its ability to understand the changing patterns of data.

F. Ablation Experiments

As shown in Fig. 4, a series of ablation experiments are designed in this paper in order to deeply understand the contribution of each component of the model to the prediction performance. Starting from the basic single-layer LSTM model, we gradually add components such as double-layer LSTM structure, self-attention mechanism, multi-head attention, residual connection, and layer normalization, and finally construct the complete LT-Hybrid model, and systematically analyze the roles of each module.

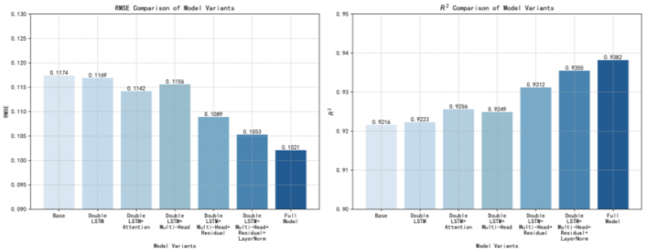

Fig. 4. Ablation experiment.

The experimental results show that the base single-layer LSTM model (Base) exhibits basic timing modeling capabilities, achieving an RMSE of 0.1174 and an R^2 value of 0.9216. On this basis, a slight improvement in model performance is obtained after upgrading to a two-layer LSTM structure (RMSE decreases to 0.1169 and R^2 improves to 0.9223), indicating that simply increasing the depth of the network does not significantly improve prediction. The introduction of the self-attention mechanism showed a significant improvement in model performance (RMSE decreased to 0.1142 and R^2 improved to 0.9256), which validates the effectiveness of the attention mechanism in capturing temporal feature correlations. However, a slight fluctuation in model performance was observed when upgrading to the multi-attention structure (RMSE slightly increased to 0.1156 and R^2 slightly decreased to 0.9249), and this temporary performance fallback suggests that the improved model structure may require a more optimal parameter configuration to be effective.

Notably, a significant jump in model performance was observed after the introduction of residual connectivity (RMSE decreased to 0.1089 and R^2 improved to 0.9312), suggesting that the residual structure effectively mitigates the gradient problem in deep network training. Further addition of layer normalization improves the stability and performance of the model (RMSE drops to 0.1053 and R^2 improves to 0.9355). The final complete model achieves optimal prediction performance (RMSE of 0.1021 and R^2 of 0.9382) through the synergy of the components.

The results of the ablation experiments clearly demonstrate the importance of each model component, especially the introduction of residual linking and layer normalization plays a key role in model performance improvement. At the same time, the performance fluctuations during the experiments reflect the complexity of deep learning model optimization, and certain architectural improvements may need to be synergized with other components for maximum effect. This series of experiments verifies the reasonableness of the hybrid architecture design proposed in this paper, and also provides a valuable reference for subsequent model improvement.

G. Hyperparametric Experiments

As shown in Fig. 5, in order to deeply study the stability of the model and determine the optimal configuration, this paper conducts systematic experimental analysis on four key hyperparameters of the LT-Hybrid model, including Learning Rate, Batch Size, Number of Attention Heads, and Hidden Layer Dimension (Hidden Size).

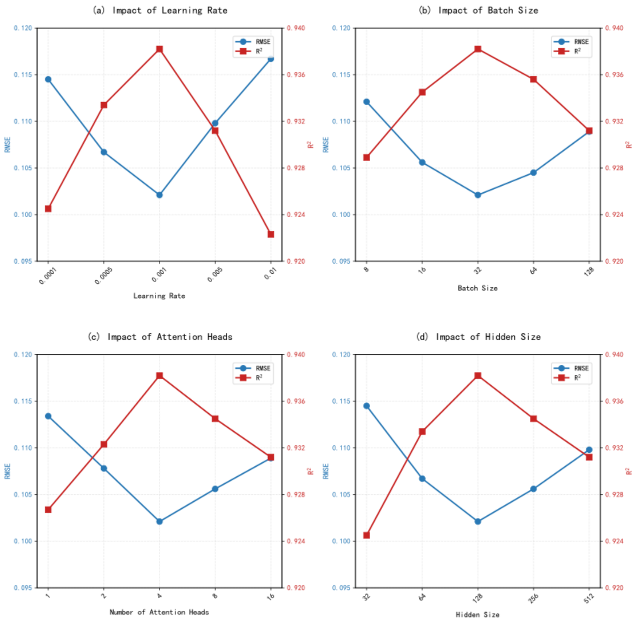

Fig. 5. Hyperparameter experiment.

In terms of learning rate, the experimental results show that 0.001 is the optimal choice, and the model obtains the lowest RMSE (0.1021) and the highest R² (0.9382) at this value point. When the learning rate is too small (e.g., 0.0001), the model converges slowly and the performance is limited; when the learning rate is too large (e.g., 0.01), the model struggles to converge stably, resulting in a significant degradation in performance. This finding is in line with the general rule of learning rate setting in deep learning, which is to ensure that the model has sufficient learning capability while avoiding too large parameter update step size.

For the choice of batch size, experiments show that 32 is the more desirable configuration. With this batch size, the model maintains a better generalization ability and also makes full use of GPU resources. It is worth noting that when the batch size is too small (8 or 16), the model training is not stable enough; while when the batch size is too large (64 or 128), although the training process is smoother, the model's performance shows a slight degradation, which may be due to the fact that the large batch training reduces the model's generalization ability.

In terms of the configuration of the attention mechanism, setting up four attention heads can achieve optimal results. The experimental results show that a single attention head performs relatively poorly (RMSE of 0.1134), which indicates that a single attention mechanism is difficult to adequately capture feature associations on different time scales. As the number of attentional heads increases, the model performance first improves and then decreases, which indicates that too many attentional heads may introduce redundant information and affect the prediction accuracy of the model instead.

Experiments on hidden layer dimensions suggest that 128 is the most appropriate choice. Smaller hidden layer dimensions (e.g., 32) limit the expressive power of the model, while too large dimensions (e.g., 512) may lead to overfitting and can significantly increase the computational overhead. With a dimension of 128, the model achieves a good balance between expressiveness and computational efficiency.

Through this series of hyper-parameter experiments, we not only determine the optimal configuration of the model, but also gain a deeper understanding of the influence mechanism of each hyper-parameter on the model performance, which provides an important reference for subsequent model optimization and application. The experimental results also verify the stability of the model under different parameter configurations, demonstrating the good generalization ability and robustness of the LT-Hybrid model.

V. Conclusion

In this paper, a hybrid deep learning architecture LT-Hybrid based on LSTM and Transformer is proposed for the air quality prediction problem. The model captures the long-term dependencies of the time series data through a two-layer LSTM structure, models the complex interactions among different environmental factors by using the multi-head self-attention mechanism, and adopts a combination of residual linkage and layer normalization which is designed to improve the training stability of the model. Experiments on the urban air quality dataset, which contains nine dimensions of environmental characteristics such as temperature, humidity, PM2.5, etc., show that the LT-Hybrid model achieves an RMSE of 0.1021 and an R² value of 0.9382, which is a significant performance enhancement compared with benchmark models such as the traditional LSTM and XGBoost. In addition, the effectiveness of each core component of the model is verified through systematic ablation experiments and hyperparameter analysis, especially the introduction of the multi-head attention mechanism and residual structure plays a key role in model performance improvement.

Although this study has achieved good results in the air quality prediction task, there are still some directions that can be improved: first, the current model mainly focuses on the prediction of single-point locations, and in the future, it can be extended to multi-site collaborative prediction, making full use of spatial information to enhance the prediction accuracy; second, the model has relatively large prediction errors when dealing with pollution events under extreme weather conditions, and the introduction of external data sources such as meteorological forecasts can be considered to enhance the model prediction capability; finally, there is still room for optimization of the model computational complexity. We can consider introducing external data sources such as meteorological forecasts to enhance the prediction ability of the model; finally, there is still room for optimizing the computational complexity of the model, and techniques such as model compression and knowledge distillation can be explored in the future to enhance the application efficiency of the model in the actual environmental monitoring system.

REFERENCES

[1] D. Iskandaryan, F. Ramos, and S. Trilles, "Air quality prediction in smart cities using machine learning technologies based on sensor data: a review. " Applied Sciences, vol. 10, no. 7, p. 2401, 2020.

[2] W. Liu, Z. Wang, X. Liu, N. Zeng, Y. Liu, and F. E. Alsaadi, "A survey of deep neural network architectures and their applications," Neurocomputing, vol. . 234, pp. 11-26, 2017.

[3] J. Ma, J. C. Cheng, Y. Ding, and J. Lin, "A temporal-spatial interpolation and extrapolation method based on geographic Long Short-Term Memory neural network for PM2.5," Journal of Cleaner Production, vol. 237, p. 117729, 2019.

[4] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," Advances in Neural Information Processing Systems, vol. 30, pp. 5998-6008, 2017.

[5] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785-794, 2016.

[6] Y. Qi, Q. Li, H. Karimian, and D. Liu, "A hybrid model for spatiotemporal forecasting of PM2.5 based on graph convolutional neural network and long short-term memory," Science of the Total Environment, vol. 664, pp. 1-10, 2019.

[7] X. Zhang, Y. Chen, J. Fan, and C. Li, "A novel deep learning approach for PM2.5 concentration forecasting based on 1D-CNN and bi-LSTM hybrid neural network," Atmospheric Pollution Research, vol. 12, no. 5, pp. 110-121, 2021.

[8] B. Zhai and J. Chen, "Development of a stacked ensemble model for forecasting and analyzing daily average PM2.5 concentrations in Beijing, China," Science of the Total Environment, vol. 635, pp. 644-658, 2018.

[9] C. Wen, S. Liu, X. Yao, L. Peng, and X. Li, "A novel spatiotemporal convolutional long short-term neural network for air pollution prediction," Science of the Total Environment, vol. 654, pp. 1091-1099, 2019.

[10] Q. Tao, F. Liu, Y. Li, and D. Sidorov, "Air pollution forecasting using a deep learning model based on 1D convnets and bidirectional GRU," IEEE Access," IEEE Access. vol. 7, pp. 76690-76698, 2019.

[11] C. Huang and K. Kuo, "A deep CNN-LSTM model for particulate matter (PM2.5) forecasting in smart cities," Sensors, vol. 18, no. 7, p. 2220, 2018.

[12] X. Yi, J. Zhang, Z. Wang, T. Li, and Y. Zheng, "Deep distributed fusion network for air quality prediction," Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 965-973, 2018.

[13] B. S. Freeman, G. Taylor, B. Gharabaghi, and J. Thé, "Forecasting air quality time series using deep learning," Journal of the Air & Waste Management Association, vol. 68, no. 8, pp. 866-886, 2018.

[14] Y. Liang, S. Ke, J. Zhang, et al. "GeoMAN: Multi-level attention networks for geo-sensory time series prediction," Proceedings of the 27th International Joint Conference on Artificial Intelligence, pp. 3428-3434, 2018.

[15] B. Yu, H. Yin, and Z. Zhu, "Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting," Proceedings of the 27th International Joint Conference on Artificial Intelligence, pp. 3634-3640, 2018.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] R. Zhao, R. Yan, J. Chen, K. Mao, P. Wang, and R. X. Gao, "Deep learning and its applications to machine health monitoring," Mechanical Systems and Signal Processing, vol. 115, pp. 213-237, 2019.

[17] Z. Qi, T. Wang, G. Song, W. Hu, X. Li, and Z. Zhang, "Deep air learning: interpolation, prediction, and feature analysis of fine-grained air quality," IEEE Transactions on Knowledge and Data Engineering, vol. 30, no. 12, pp. 2285-2297, 2018.

[18] X. Zhou, J. Wang, J. Wang, and Q. Guan, "Predicting air quality using a multi-scale spatiotemporal graph attention network," Information Sciences, vol. 2024.

[19] H. Xia, X. Chen, Z. Wang, X. Chen, and F. Dong, "A Multi-Modal Deep-Learning Air Quality Prediction Method Based on Multi-Station Time-Series Data and Remote-Sensing Images: Case Study of Beijing and Tianjin," Entropy, 2024.

[20] Y. Huang, J.J.C. Ying, and V.S. Tseng, "Spatio-attention embedded recurrent neural network for air quality prediction," Knowledge-Based Systems , vol. 212, 106597, 2021.

[21] Basha, S. A. K., Vincent, P. D. R., Mohammad, S. I., Vasudevan, A., Soon, E. E. H., Shambour, Q., & Alshurideh, M. T. (2025). Exploring Deep Learning Methods for Audio Speech Emotion Detection: An Ensemble MFCCs, CNNs and LSTM. Appl. Math, 19(1), 75-85.