


**Please cite the Published Version**

Cai, J, Liu, T, Wang, T, Feng, H, Fang, K, Bashir, AK  and Wang, W (2024) Multi-Source Fusion Enhanced Power-Efficient Sustainable Computing for Air Quality Monitoring. IEEE Internet of Things Journal. p. 1.

**DOI:** <https://doi.org/10.1109/JIOT.2024.3420956>

**Publisher:** Institute of Electrical and Electronics Engineers (IEEE)

**Version:** Accepted Version

**Downloaded from:** <https://e-space.mmu.ac.uk/635861/>

**Usage rights:**  In Copyright

**Additional Information:** This is an author accepted manuscript of an article published in IEEE Internet of Things Journal. © 2024 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

**Enquiries:**

If you have questions about this document, contact [openresearch@mmu.ac.uk](mailto:openresearch@mmu.ac.uk). Please include the URL of the record in e-space. If you believe that your, or a third party's rights have been compromised through this document please see our Take Down policy (available from <https://www.mmu.ac.uk/library/using-the-library/policies-and-guidelines>)

# Multi-source Fusion Enhanced Power-efficient Sustainable Computing for Air Quality Monitoring

Jijing Cai, Tongcun Liu, Tingting Wang, Hailin Feng, Kai Fang, Ali Kashif Bashir, and Wei Wang

**Abstract**—Given the severity of air pollution, air quality monitoring has become a crucial aspect of Artificial Intelligence of Things (AIoT) applications, providing essential information for forecasting air pollution. However, the training process for air quality monitoring models heavily relies on high-performance computing resources, leading to significant energy consumption and associated carbon emissions. This contradicts the objectives of low-carbon and sustainable computing. This paper proposes a New Hybrid PM2.5 Prediction Model (NHPPM) for air quality monitoring to address the above challenges. NHPPM prioritizes energy efficiency while maintaining high prediction accuracy by integrating several power-efficient strategies. Firstly, Wiener filtering is used to denoise multi-source air quality data, enhancing the efficiency of multi-source data fusion. Secondly, Variational Mode Decomposition (VMD) decomposes different components of multi-source air quality data, helping to identify and separate the most important factors affecting pollutants. This reduces the data needed for model training and leads to lower resource consumption. Kernel Principal Component Analysis (KPCA) transforms high-dimensional data into a lower-dimensional representation while retaining critical information, further minimizing computational demands. Additionally, this paper utilizes the Informer deep learning model to analyze trends in air quality data. The model's effectiveness is validated through ablation studies, performance evaluation experiments, and short- and long-term prediction experiments. The experimental results show that our model reduces the Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) by 16.2% and 14.9%, respectively, compared to existing PM2.5 prediction models. Furthermore, it reduces the energy consumption of model training by 33.8%.

**Index Terms**—AIoT, Multi-source data fusion, Power-efficient sustainable computing, Low-carbon, Air quality monitoring

## I. INTRODUCTION

**A**CHIEVING carbon neutrality is pivotal for fostering a sustainable and low-carbon future, significantly reducing resource consumption and combating global climate change [1], [2]. By striving for a balance between greenhouse gas emissions and the Earth's capacity to absorb them, carbon

neutrality aims to limit global temperature rises, thereby diminishing the effects of climate change. This ambitious goal necessitates the gradual elimination of fossil fuels and encourages the adoption of clean, renewable energy sources, such as solar and wind power, marking a decisive step towards low-carbon, sustainable energy systems [3], [4]. Such a transition not only minimizes reliance on finite natural resources but also enhances energy efficiency, embodying the principles of sustainability. Furthermore, achieving carbon neutrality is instrumental in steering the economy towards green, low-carbon pathways. This shift promises the creation of new growth sectors and job opportunities while decreasing resource consumption and environmental degradation during energy production and consumption. It is a strategy that aligns economic progress with sustainability. Importantly, the pursuit of carbon neutrality has profound implications for public health, directly linking to sustainable living by curbing emissions of harmful air pollutants, improving air quality, and consequently reducing health hazards associated with air pollution [5], [6]. The drive towards carbon neutrality and sustainability is not just about mitigating the impact of climate change; it's a comprehensive approach that encompasses reducing resource use, safeguarding the environment, fostering economic resilience, and enhancing quality of life. Carbon neutrality stands as a critical milestone in global efforts to address climate change and paves the way for a more sustainable, equitable, and healthier future.

AIoT systems merge cutting-edge artificial intelligence technologies with an extensive Internet of Things (IoT) infrastructure, enabling real-time monitoring and data collection of the environment, devices, and user behavior through the deployment of thousands of sensors and the development of intricate communication networks [7]. This integration requires significant computational power and an uninterrupted power supply to facilitate the high-speed operations of data processing, model training, and real-time analysis. However, such reliance on high-performance computing resources leads to substantial energy consumption and increased carbon emissions. Data centers, pivotal in maintaining the operation of servers, data storage, and cooling systems, consume a large amount of electricity, especially pronounced when processing vast volumes of IoT data and executing complex AI algorithms, thus exponentially increasing their energy demands [8]–[10]. As AIoT systems become more prevalent, the electricity consumed by connected devices and sensors significantly adds to the system's overall energy requirements. Hence, integrating low-carbon and sustainable computing approaches within AIoT systems is essential. By doing so, we not only harness these

Jijing Cai, Tongcun Liu, Hailin Feng, and Kai Fang are with the College of Mathematics and Computer Science, Zhejiang A&F University, Hangzhou 311300, China. (Email: Jijingcai19@gmail.com, tongcun.liu@gmail.com, hlifeng@zafu.edu.cn, and Kaifang@ieee.org)

Tingting Wang is with the School of Computer Science and Engineering, Faculty of Innovation Engineering, Macau University of Science and Technology, Macau SAR 999078, China. (Email: tingtingwang@ieee.org)

Ali Kashif Bashir is with the Department of Computing and Mathematics, Manchester Metropolitan University, UK (Email: dr.alikashif.b@ieee.org)

Wei Wang is with the Guangdong-Hong Kong-Macao Joint Laboratory for Emotional Intelligence and Pervasive Computing, Artificial Intelligence Research Institute, Shenzhen MSU-BIT University, Shenzhen, 518172, Guangdong, China (Email: ehomewang@ieee.org)

This paper Jijing Cai and Tongcun Liu share the same contributions  
Corresponding authors: Hailin Feng and Kai Fang

systems' potential for intelligent management and decision-making but also address the challenges of achieving carbon neutrality and protecting the environment. Incorporating energy-efficient algorithms, optimizing data processing, and leveraging renewable energy sources are vital steps toward reducing the carbon footprint of AIoT systems, ensuring their operation aligns with the principles of sustainability and environmental stewardship.

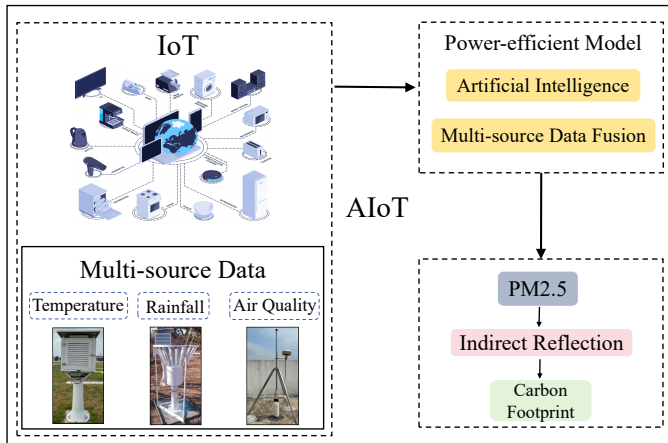


Fig. 1. Power-efficient PM2.5 Prediction Model with AIoT

To advance sustainability in AIoT systems, focusing on low-carbon, power-efficient computing is essential. By adopting efficient technologies and streamlining data workflows, we can minimize energy use and carbon emissions, aligning AIoT development with environmental goals. Rong G et al. [11] introduced an industrial-grade edge-cloud collaborative computing platform, Sophon Edge, which helps efficiently build and deploy AIoT applications. As an enterprise-level solution for the AIoT computing paradigm, Sophon Edge adopts a pipeline-based computing model for streaming data from IoT devices. Furthermore, the platform supports iterative model evolution and updates, making AIoT applications agile and data-driven. Yang et al. [12] introduced a service platform named AIoTtalk, which is the first AIoT service platform based on the SIP protocol, aimed at rapidly developing scalar and multimedia AIoT applications. Unlike commonly used IoT protocols (such as MQTT or CoAP), SIP is more suitable for complex AIoT applications. AIoTtalk also includes an experimental test platform and two SIP-based AIoT application examples, demonstrating its effectiveness in cloud and edge computing environments. Experimental results show that AIoTtalk can achieve low latency and high-quality AIoT application experiences. The aforementioned studies mainly focus on improving the performance of AIoT but do not delve into the potential applications of AIoT in reducing carbon emissions.

This paper introduces an innovative and power-efficient air quality monitoring model based on AIoT technology. The specific input, processing, and output process are illustrated in Fig. 1. Initially, the IoT system provides multi-source data, which is then analyzed and processed for prediction using

artificial intelligence technology. The model employs two techniques to achieve effective fusion of multi-source data: VMD and KPCA. VMD is applied independently to each data source, decomposing its modal functions. This step unifies the characteristics of different data sources, enhancing data consistency and compatibility, and laying a solid foundation for subsequent fusion and analysis. Next, KPCA projects these optimized data into the same high-dimensional space, extracting key features while also achieving feature dimension reduction. This effectively reduces the computational burden of training the deep learning model and lowers energy consumption. Finally, the Informer deep learning model is trained and outputs prediction results. The Informer model is particularly suitable for processing long-term sequential data and can effectively predict PM2.5 variations. Its advantage lies in the ability to process a large amount of input features and predict future PM2.5 conditions. Since PM2.5 and carbon emissions often originate from the same or related activities, this system, by accurately predicting PM2.5 concentrations, can indirectly reflect a region's carbon emissions. This offers an economical and effective alternative for regions lacking direct carbon emission monitoring equipment or technology.

The main contributions of this paper are as follows:

- 1) This paper introduces the fusion of multi-source air quality data using VMD and KPCA. VMD decomposes each data set into its modal functions, unifying the characteristics of different data sources, which enhances data consistency and compatibility. KPCA extracts key features from the optimized data and reduces its dimensionality, effectively lightening the computational load of training the deep learning model and reducing energy consumption.
- 2) This paper proposes a New Hybrid PM2.5 Prediction Model for PM2.5 prediction, which maintains a high prediction accuracy while consuming lower energy during the training process. Accurately predicting PM2.5 concentrations indirectly reflects a region's carbon emission situation.
- 3) This paper validates the high energy efficiency and accuracy of the model through ablation experiments, efficacy evaluation experiments, and short-term and long-term prediction experiments on two real datasets. Compared to existing PM2.5 prediction models, the model presented in this paper has reduced the MAE and RMSE in PM2.5 prediction by 16.2% and 14.9%, respectively, and has reduced the energy consumption in training by 33.8%.

The subsequent sections of this paper are organized as follows: In Section II, we delve into related work, examining past research on PM2.5 prediction and the utilization of deep learning in PM2.5 forecasting. Section III details the framework of our NHPPM, covering the data processing module and the Informer prediction model within the forecasting framework. Section IV details comparative experiments, providing an analysis and evaluation of the effectiveness of our proposed NHPPM. Section V is the Acknowledgement. Finally, Section VI summarizes research outcomes.

## II. RELATED WORK

In the field of PM<sub>2.5</sub> concentration prediction, conventional methodologies encompass time series analysis, regression, and physicochemical models. Time series analysis methods typically involve autoregressive and moving average models, as well as their combinations such as Autoregressive Moving Average and Autoregressive Integrated Moving Average (ARIMA) [13]. Kaur et al. [14] proposed a hybrid model that combines discrete wavelet decomposition with ARIMA to predict future PM<sub>2.5</sub> concentrations, yielding satisfactory predictive performance according to experimental results. Additionally, the grey model finds application in certain domains. Regression models, including linear regression and multiple regression, typically leverage historical data and pertinent features for PM<sub>2.5</sub> concentration prediction. Physicochemical models simulate particulate matter's generation and dispersion processes in the air to infer PM<sub>2.5</sub> concentrations. Despite offering useful predictions in specific scenarios, these approaches are often constrained by data quality and modeling assumptions, rendering them less adept at handling complex nonlinear environments.

As an extension of traditional approaches, machine learning methods have gained widespread application in predicting PM<sub>2.5</sub> concentrations. Support vector regression, decision trees, random forests, and k-nearest neighbors are commonly employed for tackling PM<sub>2.5</sub> concentration prediction challenges. Xiao et al. [15] introduced a novel machine-learning model that imputes missing satellite data through multiple estimations, followed by spatial clustering to delineate modeling regions. A set of machine learning models, including random forests, generalized additive models, and extreme gradient boosting models, were trained in each region separately. Finally, a generalized additive ensemble model is developed to integrate predictions from different algorithms. While these methods offer the advantage of automatically handling feature selection and modeling, they often require extensive tuning for optimal performance. Additionally, they may struggle to capture complex nonlinear relationships in the data, particularly in multimodal data scenarios.

Deep learning methods have spearheaded the development of PM<sub>2.5</sub> concentration prediction. Convolutional Neural Networks (CNN) [16]–[18], Recurrent Neural Networks (RNN) [19]–[21], Long Short-Term Memory (LSTM) [22], and Transformer architectures [23] are extensively employed for processing time series data. These approaches showcase remarkable performance in PM<sub>2.5</sub> concentration prediction tasks because they automatically extract features, capture complex nonlinear relationships, and exhibit strong generalization capabilities. Another strength of deep learning methods is their ability to integrate and fuse features from multiple data sources, enhancing the accuracy of PM<sub>2.5</sub> concentration predictions. Faraji et al. [24] proposed a novel PM<sub>2.5</sub> prediction model named CNN-GRU. This model, composed of a triple convolutional neural network and gated recurrent units, strongly predicts PM<sub>2.5</sub> concentrations. Du et al. [25] introduced a novel deep learning model for PM<sub>2.5</sub> prediction, comprising one-dimensional convolutional neural networks

and Bidirectional Long Short-term Memory (BiLSTM). Experimental validation demonstrates its commendable accuracy in forecasting PM<sub>2.5</sub> concentrations. M.-C et al. [26] introduced a composite neural network model for PM<sub>2.5</sub> concentration prediction. This model comprises a set of both pre-trained and uninitialized neural network models, showcasing robust predictive performance. Huang et al. [27] introduced a novel prediction model that tackles the non-stationarity of air quality data, leading to improved predictive accuracy. This model utilizes Empirical Mode Decomposition to decompose air quality data. Experimental results showcase the model's exceptional predictive performance. Qi Y et al. [28] introduced a hybrid model named GC-LSTM, utilizing deep learning methods. This model integrates graph convolutional networks and long short-term memory networks. Experimental results showcase the model's commendable performance in prediction. Yan et al. [29] introduced a spatiotemporal clustering prediction model, comparing its performance with the Back-propagation Neural Network (BPNN). Experimental analysis reveals that CNN-LSTM and LSTM generally outperform CNN and BPNN. Zhang et al. [30] proposed a novel air quality prediction model, CNN-LSTM, which utilizes CNN to extract data features and LSTM to train and obtain future air quality data. Experimental results demonstrate that the predictive accuracy of CNN-LSTM surpasses that of other individual models. Yanlai et al. [31] introduced a novel air quality prediction model, Deep Multi-output LSTM (DM-LSTM), which incorporates batch gradient descent, random neuron dropout, and L2 regularization algorithms. The results demonstrate that the proposed DM-LSTM model improves the stability and accuracy of air quality predictions.

In the research mentioned above, PM<sub>2.5</sub> prediction models have not fully considered the issue of energy consumption during the model training process, which becomes especially critical in the face of increasingly severe environmental challenges and demands for energy efficiency. To address this issue, this paper proposes an energy-efficient PM<sub>2.5</sub> prediction model, namely NHPPM, aimed at enhancing energy efficiency. The model integrates two advanced technologies, VMD and KPCA, achieving significant energy consumption reductions while demonstrating advantages in improving predictive performance. Specifically, the application of VMD technology within the model aims to optimize the characteristics of the original air quality data. By decomposing the air quality data into a series of modal functions with different frequency characteristics, VMD effectively reduces the complexity and randomness of the data. This improvement in data characteristics helps enhance the model's performance during the training phase and increases the accuracy of the model's predictions for future PM<sub>2.5</sub> concentration changes. Moreover, the introduction of KPCA further enhances the model's processing capability. By mapping the VMD-optimized data into a high-dimensional feature space using kernel techniques, KPCA can effectively identify and extract nonlinear features from the data. Furthermore, by performing dimensionality reduction on the high-dimensional feature space, KPCA not only significantly reduces the volume of data the model needs to process but also retains the most critical information, thereby



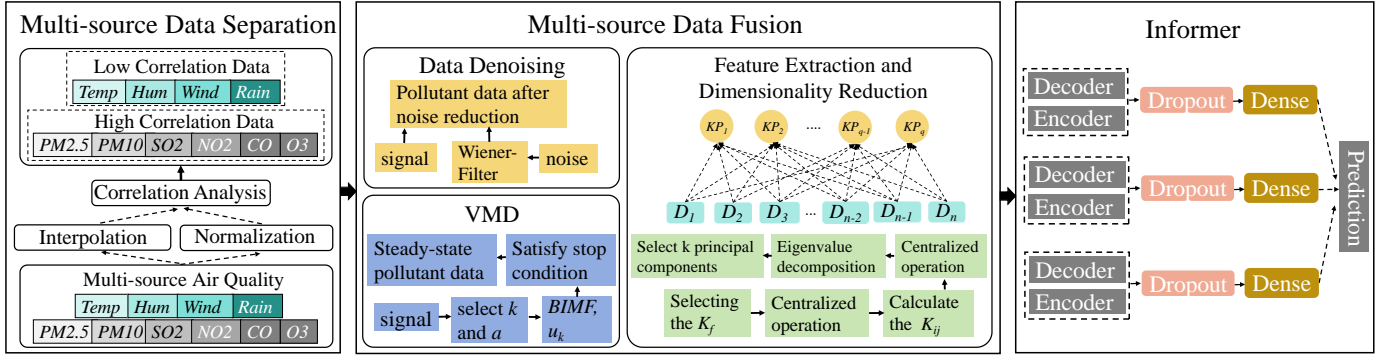


Fig. 2. Power-efficient PM2.5 Forecasting Framework with AIoT

substantially reducing the model's computational burden and energy consumption without sacrificing prediction accuracy.

### III. METHODOLOGY

The overall framework of the PM2.5 prediction model proposed in this paper is illustrated in Figure 2, which is divided into three main parts. The first part involves the partitioning of multi-source air quality data. In this stage, the data are subjected to interpolation and normalization. Interpolation fills in spatial data gaps to ensure continuity. Normalization places the data on a common scale, eliminating the influence of units of measurement. A correlation analysis is then performed to divide the multi-source air quality data into groups of high-correlation pollutant data and low-correlation meteorological data. The second part focuses on the fusion of multi-source data. This paper uses Wiener filtering for noise reduction and integrates VMD and KPCA techniques to effectively combine the de-noised multi-source data and reduce the training energy consumption of the model. VMD is applied to each data source individually, decomposing its modal functions and unifying the characteristics of different data sources. KPCA projects these optimized data into the same high-dimensional space, extracting key features and achieving feature dimension reduction. KPCA's dimensionality reduction process significantly reduces the model's resource consumption by directly reducing the volume of data to be processed. KPCA improves data processing efficiency by identifying and retaining the most representative features in the data and eliminating redundant or less relevant information. After dimensionality reduction, the data requires fewer parameter adjustments, has less computational complexity, and directly reduces the resources needed to train the model. The simplified data speeds up the entire training process with faster convergence rates, meaning that significantly fewer computational resources are required to achieve the desired accuracy. Finally, the fused data is input into the Informer model for training and prediction. As a deep learning model designed based on the Transformer structure, the Informer effectively captures long-term dependencies through its unique sparse self-attention mechanism, optimizing computational efficiency to handle large-scale time series data while reducing resource requirements during model training.

#### A. Multi-source Data Separation

1) *Interpolation and Normalization Processing*: The PM2.5 dataset used in this paper consists of the Quzhou City PM2.5 dataset and the Beijing PM2.5 dataset [32], with detailed information on both datasets as shown in Table 1. The Bei-

TABLE I  
EXPERIMENTS DATASETS DESCRIPTION

Dataset name	Beijing PM2.5 Dataset	Quzhou PM2.5 Dataset
Data type	multivariable time series	multivariable time series
location	Beijing	Quzhou
Time	01/03/2013-28/02/2017	01/01/2016-30/12/2019
Variable number	10	10
Recorded data	35064	30198

jing dataset contains 35,064 data samples, and the Quzhou dataset contains 30,198 data samples. Each sample includes 10 data features, encompassing six types of air quality features: PM2.5, PM10, SO2, CO, NO2, and O3, along with four meteorological features: air temperature, humidity, rainfall, and wind speed. The air quality data samples are represented as  $X = [X_{PM2.5}, X_{PM10}, \dots, X_{Tem}, X_{Hum}]$ , where  $X_{PM2.5}$  represents the PM2.5 concentration,  $X_{PM10}$  represents the PM10 concentration,  $X_{Tem}$  represents air temperature, and  $X_{Hum}$  represents humidity. To ensure the completeness of the air quality data sample  $X$ , it is necessary to perform interpolation on  $X$ , with the related formula for the interpolation process as follows:

$$X_P = \frac{\sum_{i=1}^n X_i}{n} \quad (1)$$

where  $X_i$  represents the non-missing data points in the air quality data sample  $X$ ,  $n$  represents the total number of non-missing data points, and  $X_P$  represents the complete air quality data sample after interpolation. Subsequently, the complete air quality data samples are normalized, scaling the data to a specific range (e.g., between 0 and 1), which can reduce errors caused by large variations in the range of data

variables, and enhance the stability and accuracy of the data. The formula related to normalization is as follows:

$$X_M = \frac{X_P - \min(X_P)}{\max(X_P) - \min(X_P)} \quad (2)$$

where  $X_P$  is the complete air quality data after interpolation,  $\min(X_P)$  represents the minimum value in  $X_P$ ,  $\max(X_P)$  denotes the maximum value in  $X_P$ , and  $X_M$  is the air quality data after normalization.

2) *Correlation Analysis*: Performing correlation analysis on the normalized air quality data  $X_M$  can identify the features that have the most significant influence on PM2.5 prediction, helping to reduce unnecessary computational complexity and improve the efficiency and accuracy of the model. In this paper, the correlation between each feature in  $X_M$  and PM2.5 is represented by the correlation coefficient  $r$ , with the formula for the correlation coefficient  $r$  as follows:

$$r = \frac{n \sum (xy) - \sum x \sum y}{\sqrt{[n \sum x^2 - (\sum x)^2][n \sum y^2 - (\sum y)^2]}} \quad (3)$$

where  $r$  represents the correlation coefficient,  $n$  denotes the sample size,  $x$  denotes the observations of the first variable, and  $y$  represents the observations of the second variable. A value of  $r$  close to 1 suggests a strong correlation between the two variables, indicating a positive trend as they change together. Table II shows the correlation coefficients between PM2.5 and air pollutant features, and Table III shows the correlation coefficient results between PM2.5 and meteorological features. The analysis shows a strong correlation between PM2.5 and PM10, CO, NO2, SO2, and O3. PM2.5 is negatively correlated with air temperature, humidity, and wind speed, and has a weak positive correlation with precipitation. Therefore, this paper divides data  $X_M$  into pollutant data  $X_{MH}$ , which is highly correlated with PM2.5, and meteorological data  $X_{ML}$ , which is lowly correlated with PM2.5. In the process of training the prediction model, pollutant data  $X_{MH}$  is used as the main input, with meteorological data  $X_{ML}$  as the auxiliary input.

TABLE II  
CORRELATION COEFFICIENT OF POLLUTANT FACTORS

Pollutant factors	PM10	SO2	CO	NO2	O3
Correlation coefficient	0.817	0.267	0.783	0.383	0.127

TABLE III  
CORRELATION COEFFICIENT OF METEOROLOGICAL FACTORS

Meteorological factors	Temperatures	Humidity	Wind Speed	Rainfall
Correlation coefficient	-0.305	-0.167	-0.141	0.042

## B. Multi-source Data Fusion

1) *Wiener filter*: After the interpolation, normalization, and correlation analysis, the pollutant data  $X_{MH}$  and meteorological data  $X_{ML}$  still retain some level of noise. To improve the fusion effect between pollutant data  $X_{MH}$  and meteorological

data  $X_{ML}$ , this paper introduces Wiener filtering to denoise the data. Wiener filtering is based on statistics and the Minimum Mean Square Error criterion, aiming to find the optimal filter settings to minimize the mean square error between the original signal and the estimated signal [33]. Its core idea is to design the filter using the statistical characteristics of the signal and noise. Fig. 3 illustrates the principle of the Wiener filter. The input is a noisy signal denoted as  $Y_M^*$ , and the output after filtering is the denoised signal denoted as  $Y_M$ . The denoising process involves the following steps: first, determine the power spectra of the signal and noise, where the signal power spectrum  $S_{XX}(f)$  is based on prior knowledge of the signal or calculated from clean signal samples, and the noise power spectrum  $S_{VV}(f)$  is estimated from the data, with  $f$  representing frequency. Then, calculate the transfer function  $H(f)$  of the filter, with the related formula as follows:

$$H(f) = \frac{S_{XX}(f)}{S_{XX}(f) + S_{VV}(f)} \quad (4)$$

The input data  $X_M^*$  in the time domain is transformed to the

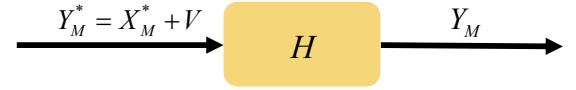


Fig. 3. Wiener Filter Structure Diagram.

frequency domain  $X_M^*(f)$  using the Fourier transform, where  $X_M^* = \{X_{MH}^*, X_{ML}^*\}$ . It is assumed that the data  $X_M^*$  and the noise are additive, such that  $Y_M^*(f) = X_M^*(f) + N(f)$ , where  $N(f)$  denotes the noise. In the frequency domain, the filtered data can be obtained by applying the filter transfer function  $H(f)$  to  $Y_M^*(f)$ , as expressed by the following formula:

$$Y_M(f) = H(f) \cdot Y_M^*(f) \quad (5)$$

where  $Y_M(f)$  is the signal after filtering. Finally,  $Y_M(f)$  is transformed back to the time domain through the inverse Fourier transform, obtaining the denoised data  $Y_M = \{Y_{MH}, Y_{ML}\}$ , where  $Y_{MH}$  represents the denoised pollutant data, and  $Y_{ML}$  represents the denoised meteorological data.

2) *Variational Modal Decomposition*: After noise reduction, the pollutant data  $Y_{MH}$  is decomposed into several modal components by VMD. Each modal component has its own central frequency and bandwidth, enabling it to adapt to signal characteristics across different frequencies and timescales [34]. VMD helps identify and extract key features within the pollutant data. These features can be used for subsequent air quality data fusion and analysis to improve the efficiency of data fusion. The flowchart for the VMD decomposition of pollutant data  $Y_{MH}$  is shown in Fig. 4.

The fundamental concept of VMD involves formulating and solving a variational problem, expressed mathematically as follows:

$$\left\{ \begin{array}{l} \min_{\{n_l\}, \{m_l\}} \left\{ \sum_l \|\partial_t [(\delta(t) + p/\pi t) * n_l(t)] e^{-pm_l t}\|^2 \right\} \\ s.t. \sum_l n_l = h \end{array} \right. \quad (6)$$

where  $n_l$  and  $m_l$  represent the decomposed  $l$ -th modal component and center frequency, respectively.

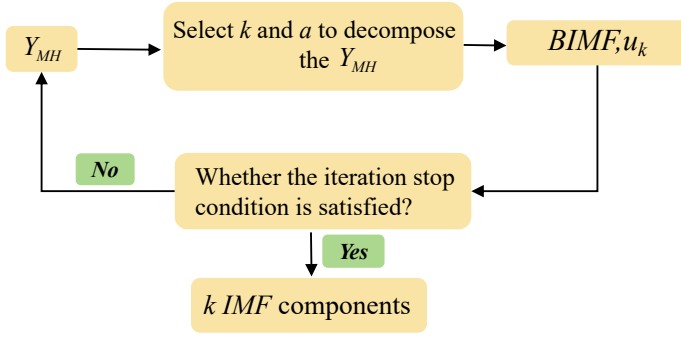


Fig. 4. Block Diagram of Variational Modal Decomposition.

To transform the variational problem and derive a non-linear expression, the augmented Lagrange expression is obtained as follows:

$$L(\{n_l\}, \{m_l\}, \sigma) = \alpha \sum_l \|\partial_t [(\delta(t) + p/\pi t) * n_l(t)] e^{-pm_l t}\|^2 + \left\| h(t) - \sum_l n_l(t) \right\|^2 + \left\langle \sigma(t), h(t) - \sum_l n_l(t) \right\rangle \quad (7)$$

where  $\alpha$  represents the penalty parameter, and  $\sigma$  is the multiplication operator.

Subsequently, utilizing the Alternating Direction Method of Multipliers optimization, the expressions for each modal component and center frequency after alternating optimization iterations  $n_l$ ,  $m_l$ , and  $\lambda$  are as follows:

$$\hat{n}_l^{n+1}(m) = \frac{\hat{h}(m) - \sum_{i \neq l} \hat{n}_i(m) + \hat{\sigma}(m)/2}{1 + 2\alpha(m - m_k)^2} \quad (8)$$

$$m_k^{n+1} = \frac{\int_0^\infty m |\hat{n}_l^{n+1}(m)|^2 dm}{\int_0^\infty |\hat{n}_l^{n+1}(m)|^2 dm} \quad (9)$$

$$\hat{\sigma}^{n+1}(m) = \hat{\sigma}^n(m) + \gamma \left( \hat{h}(m) - \sum_l \hat{n}_l^{n+1}(m) \right) \quad (10)$$

where  $\hat{h}(m)$ ,  $\hat{n}_i(m)$ , and  $\hat{\sigma}_i(m)$  denote the Fourier transforms of their respective variables.

In this paper, the VMD function is denoted as  $P_V$ . Applying  $P_V$  to the denoised pollutant data  $Y_{MH}$  yields the modal components and center frequency data of the pollutant data, denoted as  $Y_{MV}$ . The process is described by the mathematical expression as follows:

$$Y_{MV} \leftarrow P_V(Y_{MH}) \quad (11)$$

3) *Kernel Principal Component Analysis*: In this paper, we represent the KPCA method as  $P_K$ . After VMD decomposition, the data  $X_{WV}$  is processed through KPCA to obtain  $X_K$ . This process can be described using the following formula:

$$X_K \leftarrow P_K(X_{WV}) \quad (12)$$

The algorithmic steps of KPCA are depicted in Algorithm 1.

#### Algorithm 1 KPCA Algorithmic

**Input:** VMD Decomposed Data  $Y_{MV}$  and Meteorological Data  $Y_{ML}$

**Output:** Fused Data  $Y_K$

Selecting a suitable kernel function  $K_f$  for mapping the  $Y_{ML}$  and  $Y_K$  into a high-dimensional feature space.

**for** each  $Y_{ML}$  and  $Y_K$  **do**

Centering the data is done to ensure a zero mean. The centralized data is denoted as  $\varphi(x_i)$ , where  $x_i$  represents the original data point.

**end for**

Computing the inner product (kernel matrix) of the data in the high-dimensional feature space, where the elements  $K_{ij}$  in the kernel matrix  $K$  are denoted as  $K_{ij} = k(\varphi(x_i), \varphi(x_j))$

**for** kernel matrix  $K$  **do**

Performing centralization on the kernel matrix to ensure its mean is 0.

Performing eigenvalue decomposition on the centralized kernel matrix  $K$  to obtain eigenvalues  $\lambda_1, \lambda_2, \dots, \lambda_n$  and their corresponding eigenvectors  $\alpha_1, \alpha_2, \dots, \alpha_n$ .

Selecting the top  $k$  eigenvectors corresponding to the eigenvalues, where  $k$  is the desired number of principal components to retain, typically  $k < n$ .

**end for**

#### C. Informer

Input the fused data  $Y_K$  into the Informer model for Training. Informer is a supervised learning model based on attention mechanisms, consisting primarily of an encoder and a decoder. The specific structures of the encoder and decoder are illustrated in Fig. 5.

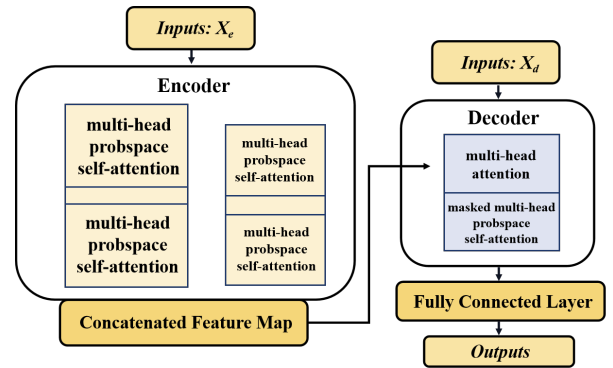


Fig. 5. Architecture of Informer Model.

The Informer has been improved upon the Transformer model. In the encoder layers, Zhou et al. [35] introduced a multi-head probability sparse self-attention mechanism to reduce the computational complexity of traditional self-attention. They employed self-attention distillation to reduce sequence length and shrink network dimensions progressively. A generative decoding mechanism was proposed in the decoder layers to directly output results during sequence prediction, thereby enhancing the model's prediction speed.

The attention mechanism primarily consists of three vector matrices: the query vector matrix Query ( $Q \in R^{L_Q \times d}$ ), the key vector matrix Key ( $K \in R^{L_K \times d}$ ), and the value vector matrix Value ( $V \in R^{L_V \times d}$ ), where  $L_Q$  represents the sequence dimension of the query vector matrix,  $L_K$  represents the sequence dimension of the key vector matrix,  $L_V$  is the sequence dimension of the value vector matrix, and  $d$  is the input dimension. The mathematical expression is as follows:

$$A(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d}}\right)V \quad (13)$$

The probability form of the attention coefficients for the  $i$ -th Query is expressed as:

$$A(q_i, K, V) = \sum_j \frac{k(q_i, k_j)}{\sum_l k(q_i, k_l)} v_j = E_{p(k_i|q_i)}[V_j] \quad (14)$$

where  $p(k_i|q_i) = \frac{k(q_i, k_j)}{\sum_l k(q_i, k_l)}$ ,  $k(q_i, k_i)$  selecting asymmetric exponent  $\exp(q_i k_j^T / \sqrt{d})$ ,  $E$  represents the expected value, and  $P$  represents the conditional probability.

To differentiate the contribution of each Query to the Value, the sparsity evaluation formula for the  $i$ -th Query is defined as:

$$M(q_i, K) = \ln \sum_{j=1}^{L_K} e^{\frac{q_i k_j^T}{\sqrt{d}}} - \frac{1}{L_K} \sum_{j=1}^{L_K} \frac{q_i k_j^T}{\sqrt{d}} \quad (15)$$

where the former is the Log-Sum-Exp (LSE) of  $q_i$  across all keys, and the latter is its arithmetic mean. Based on the above analysis, the formula for the Probability Sparse Self-Attention mechanism is derived as follows:

$$A(Q, K, V) = \text{Softmax}\left(\frac{\bar{Q}K^T}{\sqrt{d}}\right)V \quad (16)$$

where the matrix  $\bar{Q}$  is obtained through the probability sparsification of the query vector matrix  $Q$ .

The Encoder is primarily employed to capture long-range dependencies in the sequence by passing the input through the multi-head probabilistic sparse self-attention module. A distillation layer is also applied to reduce network parameters and emphasize features. The distillation principle is expressed as:

$$X_{j+1}^t = \text{MaxPool}(\text{ELU}(\text{Convld}([X_j^t]_{AB}))) \quad (17)$$

where  $\text{MaxPool}$  represents the max-pooling operation,  $\text{ELU}$  is the activation function, and  $\text{Convld}$  denotes a one-dimensional convolution operation on the sequence.

The role of the Decoder is to perform forward computations and predict long sequential outputs. Moreover, the generative structure of Informer can generate the entire prediction sequence at once, greatly enhancing the efficiency of prediction decoding. The input format for the Decoder is:

$$X_{de} = \text{Concat}(X_{token}, X_0) \in R^{(L_{token} + L_y) \times d_m} \quad (18)$$

where  $X_{token}$  represents the start character;  $X_0$  is a placeholder character;  $\text{Concat}$  indicates the concatenation and merging of  $X_{token}$  and  $X_0$ .

## IV. EXPERIMENTS

### A. Experimental Configuration

Each model's training environments were configured identically to ensure the rationality of the experimental results. The software configuration was based on the TensorFlow-powered open-source deep learning library, Keras. The hardware configuration included a CPU: Intel(R) i5-10400F 2.90GHz and GPU: NVIDIA RTX 2060 SUPER.

The configuration of hyperparameters plays a crucial role in training deep learning models. Effectively modeling deep neural networks requires the setup of numerous hyperparameters. Each model undergoes training for 100 epochs with a batch size of 16 in this context. The Adam optimizer is used for training, with a dropout rate of 0.2 and ReLU activation. The training process is guided by minimizing the Mean Squared Error loss.

### B. Experimental Models

The NHPPM model proposed in this paper is compared with the following models in our experimental evaluations, each of which is introduced as follows:

1) *RNN*: RNNs are deep learning models designed for sequential data, where each input is linked to its predecessor. Unlike traditional neural networks, RNNs have internal states, or memory, allowing them to capture temporal dependencies and sequential relationships. This makes RNNs ideal for tasks involving time-series data, such as natural language processing, speech recognition, and financial forecasting. By maintaining a memory of previous inputs, RNNs can understand context, handle varying sequence lengths, and predict future states based on past information, making them powerful tools for analyzing dynamic data.

2) *LSTM*: LSTM networks are an advanced variant of RNN, specifically engineered to overcome the challenges of vanishing or exploding gradients that traditional RNNs face with long sequence data. LSTMs are distinguished by their capability to capture and maintain long-term dependencies, making them particularly effective in areas such as time series analysis, natural language processing, and other sequence modeling tasks. This attribute allows them to excel in applications where understanding historical context is crucial.

3) *BiLSTM*: BiLSTM is an enhancement of the traditional LSTM model, incorporating both forward and backward hidden layers. This design enables BiLSTM to capture relationships within sequential data from both past and future contexts, significantly enriching its understanding and representation of the sequence. This dual-directional approach is particularly effective in handling sequence data, making BiLSTM a powerful tool for tasks where context from both directions is critical for accurate predictions.

4) *CNN-BiLSTM*: The CNN-BiLSTM model is a sophisticated deep learning architecture that adeptly combines CNNs and BiLSTM to process sequential data marked by temporal dependencies. This hybrid model leverages the strengths of both CNNs and BiLSTMs, excelling in capturing local features through CNN layers and long-term dependencies via BiLSTM layers. This dual capability enables the CNN-BiLSTM model

TABLE IV  
THE ANALYTICAL OUTCOMES FROM THE ABLATION EXPERIMENT CONDUCTED ON THE BEIJING DATASET

Models	RMSE				MAE			
	1h	6h	12h	24h	1h	6h	12h	24h
Informer	20.61	53.43	68.59	87.37	12.05	32.71	44.15	60.31
Wiener-Informer	18.29	51.87	66.07	81.58	10.16	31.18	42.09	54.38
VMD-Informer	19.13	40.28	54.34	69.37	11.73	28.04	37.69	45.26
KPCA-Informer	20.57	45.99	56.43	72.15	11.27	32.42	42.99	53.57
VMD-KPCA-Informer	17.05	35.61	50.55	65.96	8.54	21.65	28.07	36.46
<b>NHPPM</b>	<b>16.04</b>	<b>33.66</b>	<b>44.18</b>	<b>62.07</b>	<b>7.87</b>	<b>19.91</b>	<b>23.46</b>	<b>32.77</b>

to deliver exceptional performance in tasks involving complex sequence data, providing a robust solution for analyzing patterns that evolve over time.

### C. Evaluation Indicators

To better quantify and evaluate the performance of models, we employ RMSE and MAE as evaluation metrics. Lower values of RMSE and MAE indicate higher predictive accuracy of the models. They are shown in the following equations respectively:

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_r - y_{pr})^2} \quad (19)$$

$$MAE = \frac{1}{n} \sum_{t=1}^n |y_r - y_{pr}| \quad (20)$$

where  $n$  denotes the number of samples,  $y_r$  represents the true values, and  $y_{pr}$  is the predicted values.

### D. Ablation Experiment

Analysis of each data processing module is crucial for comparing the proposed overall model. This analysis should delve into each component's contributions and combinations to enhance the model's overall performance. The results of the ablation experiment analysis are shown in Table IV. From the table, Wiener filter noise reduction, VMD smoothing, and KPCA dimensionality reduction all positively contribute to PM2.5 prediction. The individual impact of VMD and KPCA surpasses Wiener filtering, and their combination enhances the model further. The combination of VMD and KPCA can improve the model. Due to the decomposition of data into multiple signals by VMD during the smoothing process, KPCA can then effectively extract and downsize features from the resulting data. Compared with the direct dimensionality reduction of the original data, the combination of VMD and KPCA can better extract effective information from the data and ensure its smoothness. Our proposed NHPPM model achieves good prediction results through noise reduction, smoothing, feature extraction, and dimensionality reduction.

### E. Model Energy Efficiency Evaluation Experiment

To validate the energy efficiency advantages of the proposed NHPPM model, this paper conducts an energy efficiency evaluation experiment using the early stopping mechanism on the PM2.5 dataset from Quzhou. The experiment compares the performance of various models by determining the number of training epochs required to terminate when predicting PM2.5 concentrations for the next 1, 3, 6, 12, 18, and 24 hours. The experimental results are shown in Fig. 6. At various time spans, the termination epochs of the RNN model are relatively high, indicating that the model requires more training epochs to reach the termination criterion. This is likely due to the vanishing gradient problem in long-term sequence prediction with RNNs. In contrast, the LSTM model generally has lower termination epochs, reflecting its advantages in handling long-term sequences, particularly in 1, 3, 6, and 12-hour predictions. The BILSTM model has the lowest termination epochs (3 epochs) for 1-hour predictions, but higher termination epochs for 3, 6, and 12-hour predictions, indicating that BILSTM have an advantage in short-term sequence prediction but requires more training for long-term sequences. The CNN-BILSTM model has moderate termination epochs for most time spans, especially in the 6-hour prediction (10 epochs), showing the potential of the combined convolutional and bidirectional LSTM model in handling complex time series data.

The NHPPM model has the lowest termination epochs for all prediction times, indicating that the combined model, which integrates Wiener filter, KPCA, VMD, and Informer deep learning, is the most efficient in predicting PM2.5 concentrations, achieving ideal results with fewer training epochs. The NHPPM model has the highest training efficiency, reaching the termination criterion with the fewest epochs, demonstrating that the combined model has better learning capability and generalization performance. The RNN model has the lowest training efficiency, requiring more epochs to reach the termination criterion.

The NHPPM model has the fastest convergence speed, indicating a shorter and more stable optimization path. LSTM follows, showing significant advantages in long-term sequences. BILSTM converges quickly for short-term sequences but performs generally in long-term sequences. CNN-BILSTM shows balanced performance across various time spans.



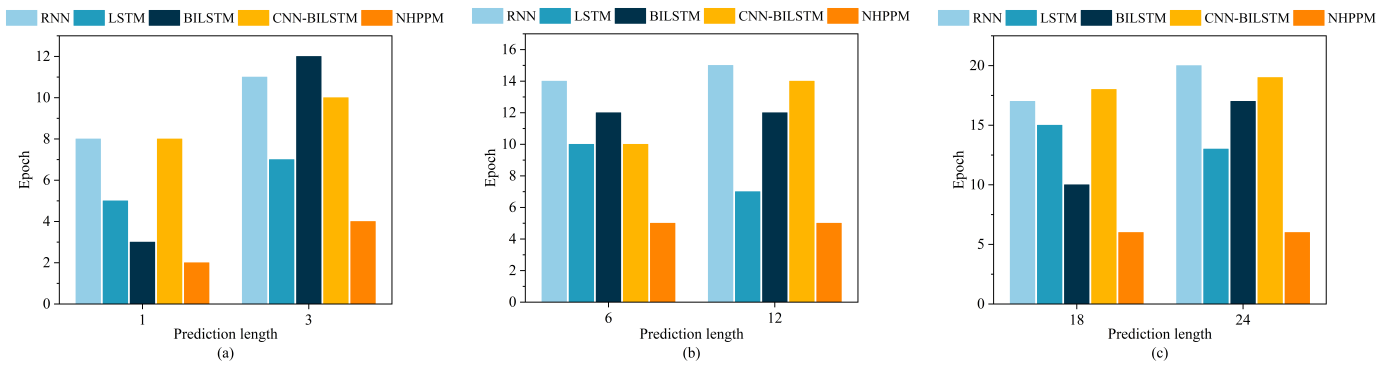


Fig. 6. Model Energy Efficiency Analysis. (a)The Number of Training Epochs to Termination for each Model Under Prediction Lengths of 1 and 3 Hours. (b)The Number of Training Epochs to Termination for each Model Under Prediction Lengths of 6 and 12 Hours. (c)The Number of Training Epochs to Termination for each Model Under Prediction Lengths of 18 and 24 Hours.

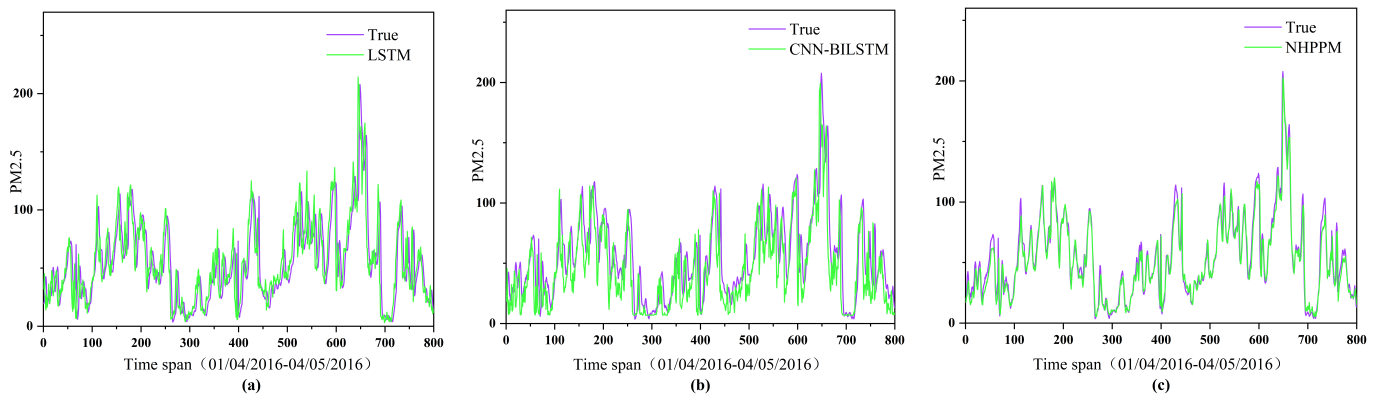


Fig. 7. In the Experiment on Beijing PM2.5 Dataset, a Comparison of Single True and Predicted PM2.5 Value During One Month (01/04/2016-04/05/2016) of Different Models (LSTM, CNN-BILSTM, and NHPPM). (a) LSTM Model. (b) CNN-BILSTM Model. (c) NHPPM Model.

In summary, the NHPPM model performs the best in predicting PM2.5 concentrations for the next 1, 3, 6, 12, 18, and 24 hours, with the highest training efficiency and fastest convergence speed. The LSTM model is suitable for long-term sequence prediction and performs well in handling complex time series data. BILSTM and CNN-BILSTM have advantages in short-term sequence prediction but require more training epochs for long-term sequences. The RNN model performs relatively poorly across all time spans, requiring more training epochs to achieve ideal results. An overall analysis of Fig. 6 shows that, compared to current mainstream prediction models, the NHPPM model can achieve optimal prediction performance in fewer training rounds, reflecting a significant energy efficiency advantage. This is due to the introduction of VMD and KPCA techniques in this paper, which enable the effective integration of multisource air quality data. The VMD technique decomposes the modal functions of each data source, unifying features and enhancing data consistency and compatibility. The KPCA technique projects these optimized data into the same high-dimensional space, extracting key features and achieving dimensionality reduction, effectively reducing the computational burden of training deep learning models and lowering energy consumption. Ultimately, by using the advanced Informer model, this paper achieves high-

energy-efficiency prediction of PM2.5 concentration.

#### F. Analysis of the Results of the Single-step Prediction Experiment

TABLE V  
THE MODEL ERROR OF NHPPM AND COMPARISONS WITH OTHER MODELS FOR THE SINGLE-STEP PM2.5 PREDICTION TASK

Models	Quzhou PM2.5 Dataset		Beijing PM2.5 Dataset	
	RMSE	MAE	RMSE	MAE
LSTM	6.58	4.54	19.06	10.80
BILSTM	6.24	4.29	19.79	11.31
RNN	6.36	4.42	19.81	11.63
CNN-BILSTM	6.03	4.12	18.51	9.97
<b>NHPPM</b>	<b>3.47</b>	<b>2.13</b>	<b>16.04</b>	<b>7.87</b>

In the single-step prediction comparative experiments, the performance metrics of each model are shown in Table V. The NHPPM model proposed in this paper excels in both the Beijing and Quzhou datasets. In the comparative experiment on the Beijing dataset, the NHPPM model achieves an MAE of 7.87 and an RSME of 16.04, significantly outperforming other deep learning models. In the comparative experiment on the Quzhou dataset, the NHPPM model achieves an MAE of



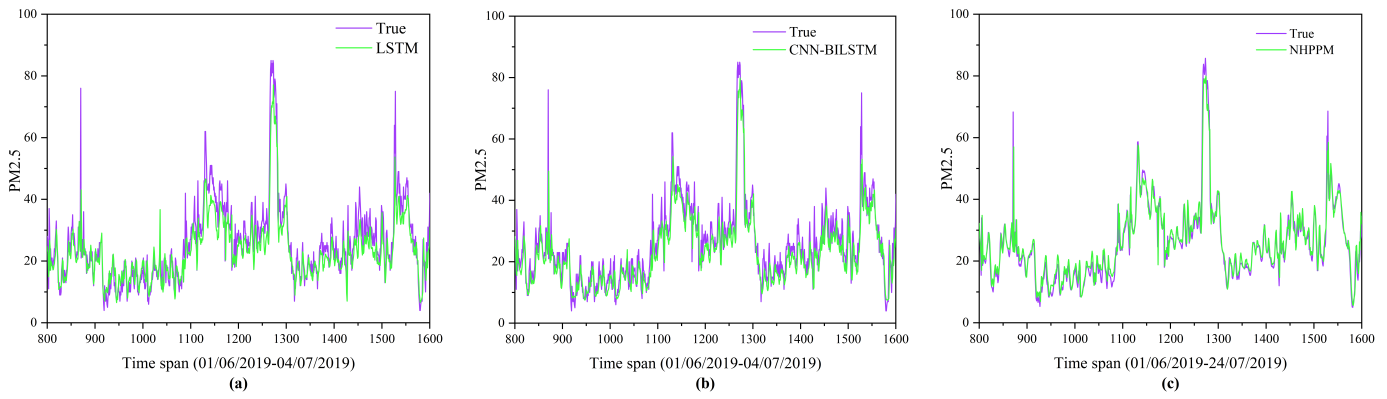


Fig. 8. In the Experiment on Quzhou PM2.5 Dataset, a Comparison of Single True and Predicted PM2.5 Value During One Month (01/06/2019-04/07/2019) of Different Models (LSTM, CNN-BILSTM, and NHPPM). (a) LSTM Model. (b) CNN-BILSTM Model. (c) NHPPM Model.

2.13 and an RMSE of 3.47. The NHPPM model outperforms its comparative models on both real-world datasets due to its comprehensive analysis and feature extraction from the air quality data, as evidenced by superior performance metrics. We first perform correlation analysis on the raw data, dividing the data into groups of highly correlated pollutant data and lowly correlated meteorological data. Additionally, we use Wiener filtering, VMD decomposition, and KPCA to denoise, extract features, and reduce the dimensionality of the pollutant data. Furthermore, we integrate the predictions from the Informer and CNN-BILSTM models on the highly correlated pollutant group. The prediction performance of CNN-BILSTM is better than other single deep learning models, demonstrating the limitations of single deep learning models in time series single-step prediction.

To further evaluate the single-step prediction performance of NHPPM and comparative deep learning models on two real-world datasets, we analyze the PM2.5 prediction capabilities of NHPPM, LSTM, and CNN-BILSTM over 800 hours of data points. Figs. 7(a), 7(b), and 7(c) in the Beijing dataset experiment compare actual and predicted values for LSTM, CNN-BILSTM, and NHPPM models. As shown in the figures, our model outperforms LSTM and CNN-BILSTM models, especially during the peak and trough periods of the air quality time series data. Figs. 8(a), 8(b), and 8(c) illustrate a comparison between actual values and single-step predicted values for LSTM, CNN-BILSTM, and NHPPM models on the Quzhou air quality dataset. These figures consistently demonstrate the superiority of the NHPPM model's single-step prediction performance compared to the LSTM and CNN-BILSTM models. In summary, our proposed model consistently performs superior in single-step prediction of air quality time series across various experimental settings. Although LSTM and CNN-BILSTM models also demonstrate good performance, the proposed model is the most effective approach.

### G. Multi-step Forecasting Results Analysis

Table VI presents the detailed results of multi-step PM2.5 forecasting for both the Beijing and Quzhou datasets, utilizing

a variety of computational models. The performance metrics, including RMSE and MAE, are calculated for forecasts extending over the next six hours. This analysis encompasses traditional models such as LSTM, BILSTM, RNN, and an advanced hybrid model, CNN-BILSTM, as well as our proposed NHPPM. As Table VI demonstrates, our NHPPM model consistently outperforms the other methods in terms of prediction accuracy for PM2.5 concentration. Specifically, our model dramatically reduces the MAE to 19.91 on the Beijing PM2.5 dataset, a notable improvement compared to existing deep learning models. Furthermore, it achieves the lowest recorded MAE of 5.52 on the Quzhou urban air quality dataset. These results not only indicate a significant advancement in prediction accuracy but also highlight the robustness and reliability of the NHPPM model across diverse environmental conditions. During the multi-step prediction experiments, we observe that the MAE and RMSE values are significantly higher for the Beijing dataset compared to those for the Quzhou dataset. This discrepancy is largely due to the distinct environmental and geographical characteristics of the two cities. Beijing, with its high population density and significant industrial activities, experiences higher levels of PM2.5 with greater temporal variability. The superior performance of our NHPPM model in both datasets suggests that it effectively captures the underlying patterns and fluctuations in PM2.5 concentrations, adapting to both the abrupt spikes seen in metropolitan Beijing and the steadier conditions in Quzhou.

TABLE VI  
THE MODEL ERROR OF NHPPM AND COMPARISONS WITH OTHER MODELS FOR THE MULTI-STEP PM2.5 PREDICTION TASK

Models	Quzhou PM2.5 Dataset		Beijing PM2.5 Dataset	
	RMSE	MAE	RMSE	MAE
LSTM	10.46	7.03	55.50	32.26
BILSTM	9.97	6.82	52.35	31.39
RNN	10.68	7.49	55.44	34.68
CNN-BILSTM	11.56	7.98	43.62	28.80
<b>NHPPM</b>	<b>8.05</b>	<b>5.52</b>	<b>33.66</b>	<b>19.91</b>

Table VII details the performance metrics for various fore-

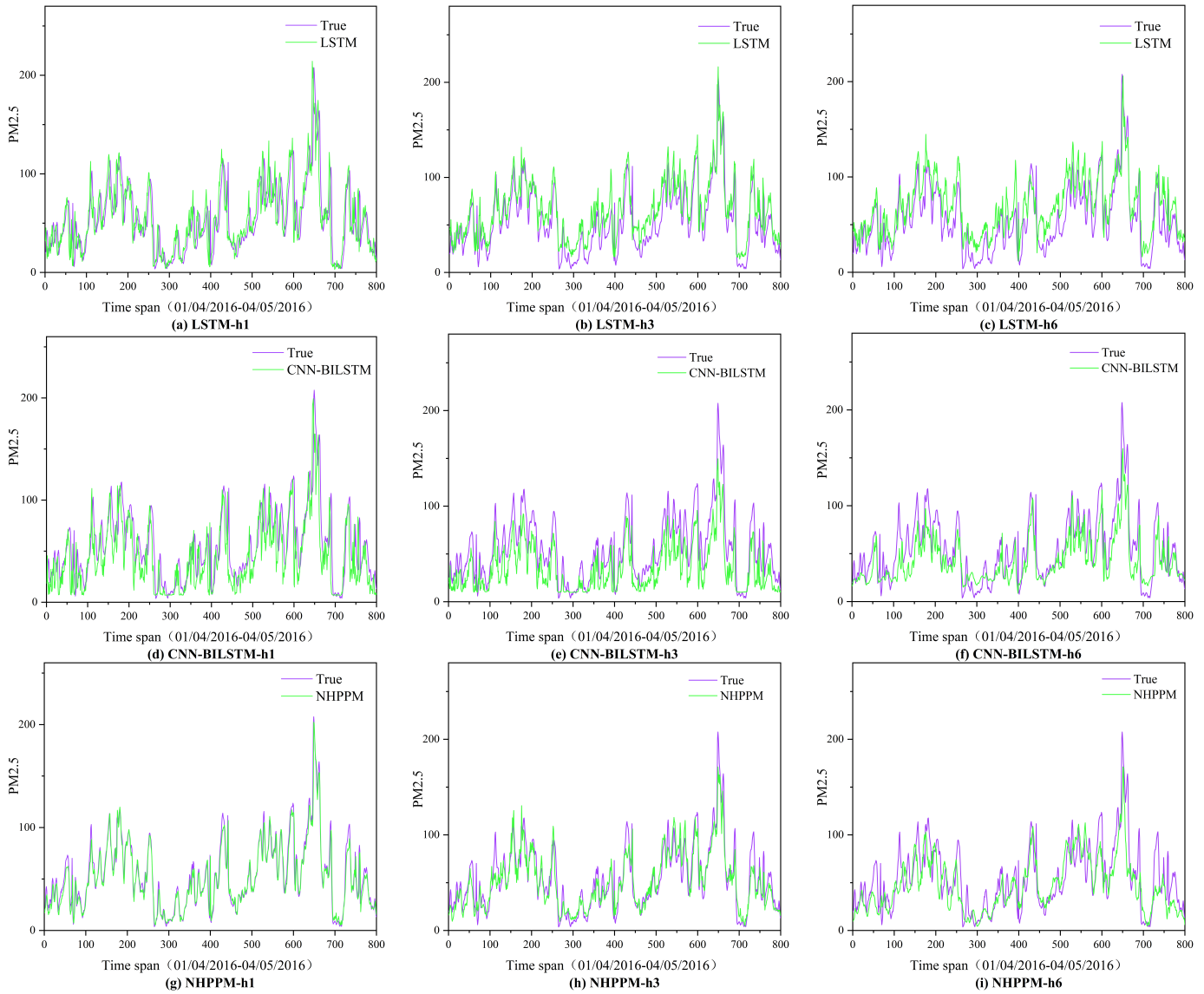


Fig. 9. In the Experiment on Beijing PM2.5 Dataset, a Comparison of Multi-step (Next h1, h3, and h6) True and Predicted PM2.5 Value During One Month (01/04/2016-04/05/2016) of Different Models (LSTM, CNN-BILSTM, and NHPPM). (a) LSTM Model for Future 1-hour Prediction. (b) LSTM Model for the Future 3-hour Prediction. (c) LSTM Model for Future 6-hour Prediction. (d) CNN-BILSTM Model for Future 1-hour Prediction. (e) CNN-BILSTM Model for the Future 3-hour Prediction. (f) CNN-BILSTM Model for the Future 6-hour Prediction. (g) NHPPM Model for Future 1-hour Prediction. (h) NHPPM Model for the Future 3-hour Prediction. (i) NHPPM Model for the Future 6-hour Prediction.

TABLE VII

IN THE EXPERIMENT ON QUZHOU PM2.5 DATASET, THE MODEL ERROR OF NHPPM AND COMPARISONS WITH OTHER BASELINE MODELS FOR THE MULTISTEP PREDICTION OF PM2.5 VALUES IN THE NEXT 24 HOURS

Models	RMSE				MAE			
	1-6h	7-12h	13-18h	19-24h	1-6h	7-12h	13-18h	19-24h
LSTM	7.77	10.35	10.71	11.33	5.39	7.32	8.08	8.56
BILSTM	7.54	10.53	11.09	11.72	5.22	7.33	7.99	8.53
RNN	7.75	10.33	10.84	11.63	5.59	7.42	8.09	8.65
CNN-BILSTM	8.07	10.51	10.92	11.17	5.75	7.41	7.83	8.07
<b>NHPPM</b>	<b>5.98</b>	<b>8.23</b>	<b>8.98</b>	<b>9.36</b>	<b>4.04</b>	<b>5.81</b>	<b>6.55</b>	<b>6.86</b>

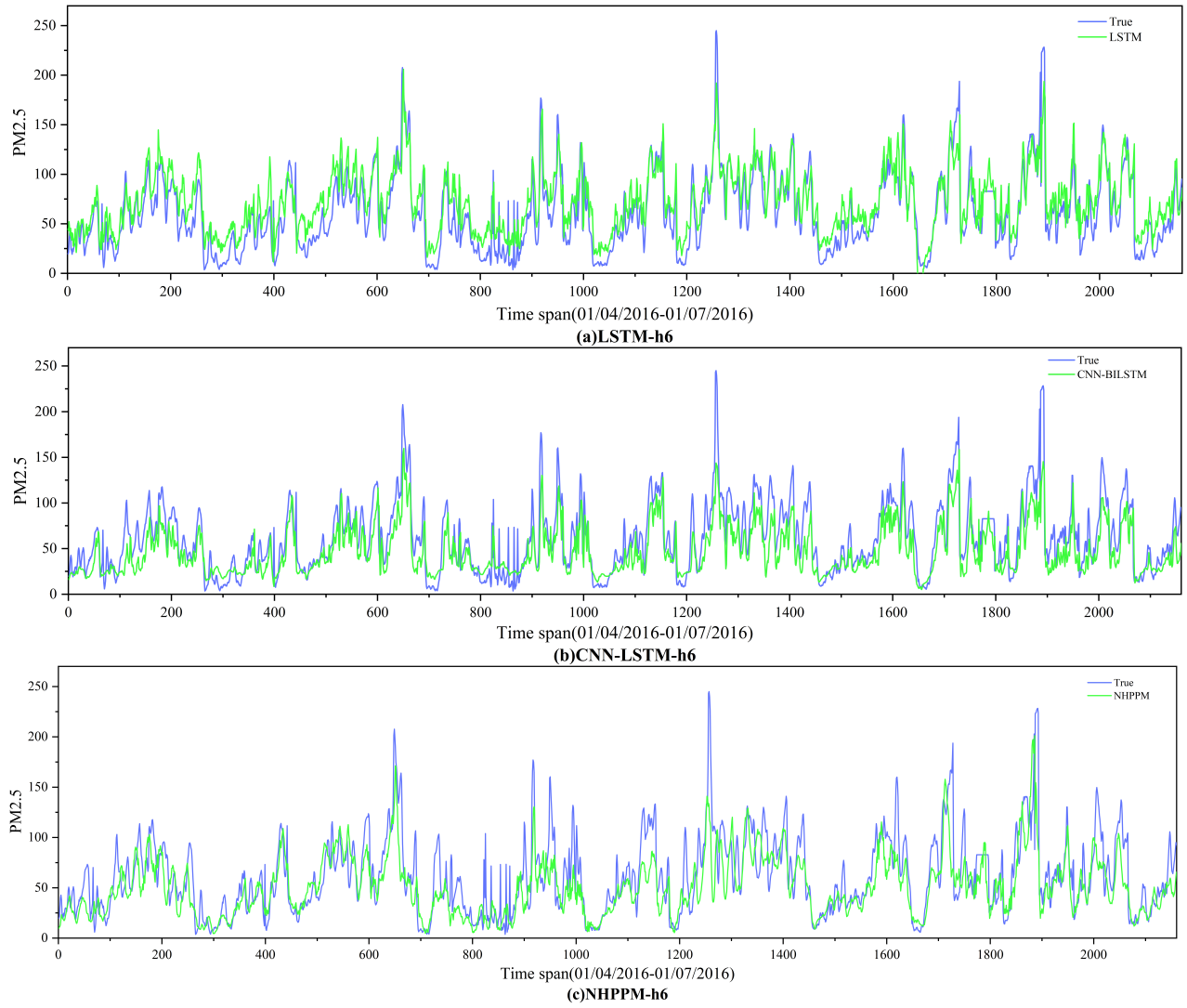


Fig. 10. In the Experiment on Beijing Dataset. A Comparison of Multi-step True and Predicted PM2.5 Value During Three Months (01/04/2016-01/07/2016) of LSTM, CNN-BILSTM, and NHPPM Model (a) LSTM for the Future 6-hour Prediction. (b) CNN-BILSTM for the Future 6-hour Prediction. (c) NHPPM for the Future 6-hour Prediction.

casting models over a 24-hour prediction horizon within the Quzhou dataset. The performance indicators are segmented into four distinct intervals: 1-6 hours, 7-12 hours, 13-18 hours, and 19-24 hours. For each interval, the average performance values are calculated, allowing for a detailed analysis across different forecasting horizons. Notably, the NHPPM model demonstrates a progressively stronger performance advantage in later intervals, with its superiority becoming more apparent in the 19-24 hour range. This pattern underscores the NHPPM model's enhanced capability in long-term forecasting while maintaining robust performance in shorter-term predictions.

To delve deeper into the comparative effectiveness of the NHPPM and traditional deep learning models, we conducted a multi-step prediction analysis using different models (LSTM, CNN-BILSTM, and NHPPM) across various step lengths (1-step, 3-step, 6-step) within the Beijing dataset, which consists of 800 data points. Fig. 9 illustrates the comparative performance, displaying the actual versus predicted values for these step lengths. The visual comparison clearly shows

that our NHPPM model surpasses both LSTM and CNN-BILSTM in multi-step forecasting. The predicted trajectories from the NHPPM model align more closely with the actual data across all tested step lengths. Moreover, as the prediction step length increases, the performance of LSTM and CNN-BILSTM models tends to deteriorate significantly, whereas the NHPPM model maintains a consistent and superior predictive accuracy.

To further analyze the predictive performance of the NHPPM model in different datasets, we conducted experiments in the Beijing dataset, dividing 1-48h into four time periods (1-6h, 7-12h, 13-24h, 25-48h). As shown in Table VIII, compared to other deep learning models, the NHPPM model has the best predictive performance in each period, proving its excellent performance in both short-term and long-term predictions.

Fig. 10 presents a comparison between the actual and predicted values of the LSTM, CNN-BILSTM, and NHPPM models in predicting the next 6 hours, based on the analysis of 2200 data points. Through the comparison of these three

TABLE VIII  
IN THE EXPERIMENT ON BEIJING DATASET, THE MODEL ERROR OF NHPPM AND COMPARISONS WITH OTHER BASELINE MODELS FOR THE MULTI-STEP PREDICTION OF PM2.5 VALUES IN THE NEXT 48 HOURS

Models	RMSE				MAE			
	1-6h	7-12h	13-24h	25-48h	1-6h	7-12h	13-24h	25-48h
LSTM	39.14	65.80	71.16	84.49	23.16	40.17	48.77	60.39
BILSTM	39.04	65.47	70.55	84.54	22.78	40.16	48.69	60.87
RNN	39.37	64.65	78.85	82.01	23.09	40.06	53.69	56.64
CNN-BILSTM	40.14	64.08	70.76	84.18	23.44	40.75	47.46	58.89
<b>NHPPM</b>	<b>31.86</b>	<b>49.40</b>	<b>65.71</b>	<b>78.12</b>	<b>18.48</b>	<b>29.11</b>	<b>40.88</b>	<b>50.74</b>

figures, it is observed that LSTM and CNN-BILSTM have limited ability to capture changes in PM2.5 concentration, especially in the peak and trough regions. In contrast, the NHPPM model exhibits better predictive ability, aligning its predicted curve more closely with the actual values.

In summary, the NHPPM model proposed in this paper has better predictive ability than traditional deep learning models. Whether in single-step or multi-step prediction, this indicates that the data processing module can effectively denoise, extract features, and reduce the dimensionality of the original air quality data.

#### V. ACKNOWLEDGMENT

This work was partly supported by the Zhejiang Provincial Natural Science Foundation of China under grant no. LQ23F020001.

#### VI. CONCLUSION

This paper proposes an innovative PM2.5 concentration prediction model aimed at reducing resource consumption during the training process. The model leverages VMD and KPCA to achieve effective multi-source air quality data fusion. VMD is applied to each data source individually to decompose its modal functions, normalizing the features of different data sources. This enhances the consistency and compatibility of the data providing a solid foundation for subsequent fusion and analysis. KPCA projects these optimized data into the same high-dimensional space, not only extracting key features but also achieving feature dimensionality reduction, effectively reducing the computational burden of training deep learning models and lowering energy consumption. Finally, by integrating the Informer model, one of the most advanced deep learning models in the time series data prediction field, the paper achieves high-accuracy PM2.5 concentration predictions with low energy consumption.

#### REFERENCES

- [1] Y. Hu and W. Wu, "Can fossil energy make a soft landing?—the carbon-neutral pathway in china accompanying ccs," *Energy Policy*, vol. 174, p. 113440, 2023.
- [2] M. Qin, C.-W. Su, O.-R. Lobont, and M. Umar, "Blockchain: a carbon-neutral facilitator or an environmental destroyer?" *International Review of Economics & Finance*, vol. 86, pp. 604–615, 2023.
- [3] N. Zeng, K. Jiang, P. Han, Z. Hausfather, J. Cao, D. Kirk-Davidoff, S. Ali, and S. Zhou, "The chinese carbon-neutral goal: challenges and prospects," *Advances in atmospheric sciences*, vol. 39, no. 8, pp. 1229–1238, 2022.
- [4] Y. Zhao, Q. Su, B. Li, Y. Zhang, X. Wang, H. Zhao, and S. Guo, "Have those countries declaring "zero carbon" or "carbon neutral" climate goals achieved carbon emissions-economic growth decoupling?" *Journal of Cleaner Production*, vol. 363, p. 132450, 2022.
- [5] A. Huovila, H. Siikavirta, C. A. Rozado, J. Rökman, P. Tuominen, S. Paiho, Å. Hedman, and P. Ylén, "Carbon-neutral cities: Critical review of theory and practice," *Journal of Cleaner Production*, vol. 341, p. 130912, 2022.
- [6] E. T. Mitchard, "The tropical forest carbon cycle and climate change," *Nature*, vol. 559, no. 7715, pp. 527–534, 2018.
- [7] H. Zhu, P. Tiwari, A. Ghoneim, and M. S. Hossain, "A collaborative ai-enabled pretrained language model for aiot domain question answering," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 5, pp. 3387–3396, 2021.
- [8] M. Shen, A. Gu, J. Kang, X. Tang, X. Lin, L. Zhu, and D. Niyato, "Blockchains for artificial intelligence of things: A comprehensive survey," *IEEE Internet of Things Journal*, 2023.
- [9] B. Yang, H. Wang, L. Hu, H. Zhu, C.-T. Lam, and K. Fang, "Few-shot cross-domain based wifi sensing system for online learning in iot," *IEEE Sensors Journal*, 2023.
- [10] T. Wang, J. Li, W. Wei, W. Wang, and K. Fang, "Deep-learning-based weak electromagnetic intrusion detection method for zero touch networks on industrial iot," *IEEE Network*, vol. 36, no. 6, pp. 236–242, 2022.
- [11] G. Rong, Y. Xu, X. Tong, and H. Fan, "An edge-cloud collaborative computing platform for building aiot applications efficiently," *Journal of Cloud computing*, vol. 10, no. 1, p. 36, 2021.
- [12] S.-R. Yang, Y.-C. Lin, P. Lin, and Y. Fang, "Aiotalk: A sip-based service platform for heterogeneous artificial intelligence of things applications," *IEEE Internet of Things Journal*, 2023.
- [13] S. Singh, A. Mohapatra *et al.*, "Repeated wavelet transform based arima model for very short-term wind speed forecasting," *Renewable energy*, vol. 136, pp. 758–768, 2019.
- [14] J. Kaur, S. Singh, and K. S. Parmar, "Forecasting of aqi (pm2. 5) for the three most polluted cities in india during covid-19 by hybrid daubechies discrete wavelet decomposition and autoregressive (db-dwd-arima) model," *Environmental Science and Pollution Research*, pp. 1–18, 2023.
- [15] Q. Xiao, H. H. Chang, G. Geng, and Y. Liu, "An ensemble machine-learning model to predict historical pm2. 5 concentrations in china from satellite data," *Environmental science & technology*, vol. 52, no. 22, pp. 13 260–13 269, 2018.
- [16] L. Yuan, H. Zhang, M. Xu, F. Zhou, and Q. Wu, "A multiscale cnn framework for wireless technique classification in internet of things," *IEEE Internet of Things Journal*, vol. 9, no. 12, pp. 10366–10367, 2022.
- [17] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980–2988.
- [18] J. Qiu, X. Yan, W. Wang, W. Wei, and K. Fang, "Skeleton-based abnormal behavior detection using secure partitioned convolutional neural



network model,” *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 12, pp. 5829–5840, 2021.

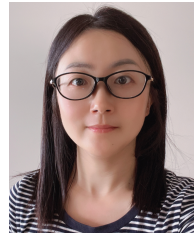
- [19] S. Chang, S. Huang, R. Zhang, Z. Feng, and L. Liu, “Multitask-learning-based deep neural network for automatic modulation classification,” *IEEE Internet of Things Journal*, vol. 9, no. 3, pp. 2192–2206, 2022.
- [20] J. Fei and L. Liu, “Real-time nonlinear model predictive control of active power filter using self-feedback recurrent fuzzy neural network estimator,” *IEEE Transactions on Industrial Electronics*, vol. 69, no. 8, pp. 8366–8376, 2021.
- [21] K. Fang, T. Wang, X. Yuan, C. Miao, Y. Pan, and J. Li, “Detection of weak electromagnetic interference attacks based on fingerprint in iiot systems,” *Future Generation Computer Systems*, vol. 126, pp. 295–304, 2022.
- [22] D. Fan, H. Sun, J. Yao, K. Zhang, X. Yan, and Z. Sun, “Well production forecasting based on arima-lstm model considering manual operations,” *Energy*, vol. 220, p. 119708, 2021.
- [23] E.-S. M. El-kenawy, F. Albalawi, S. A. Ward, S. S. Ghoneim, M. M. Eid, A. A. Abdelhamid, N. Bailek, and A. Ibrahim, “Feature selection and classification of transformer faults based on novel meta-heuristic algorithm,” *Mathematics*, vol. 10, no. 17, p. 3144, 2022.
- [24] M. Faraji, S. Nadi, O. Ghaffarpasand, S. Homayoni, and K. Downey, “An integrated 3d cnn-gru deep learning method for short-term prediction of pm2. 5 concentration in urban environment,” *Science of The Total Environment*, vol. 834, p. 155324, 2022.
- [25] S. Du, T. Li, Y. Yang, and S.-J. Horng, “Deep air quality forecasting using hybrid deep learning framework,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 33, no. 6, pp. 2412–2424, 2019.
- [26] M.-C. Yang and M. C. Chen, “Composite neural network: Theory and application to pm2. 5 prediction,” *IEEE Transactions on Knowledge and Data Engineering*, 2021.
- [27] G. Huang, X. Li, B. Zhang, and J. Ren, “Pm2. 5 concentration forecasting at surface monitoring sites using gru neural network based on empirical mode decomposition,” *Science of the Total Environment*, vol. 768, p. 144516, 2021.
- [28] 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.
- [29] 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 Systems with Applications*, vol. 169, p. 114513, 2021.
- [30] J. Zhang and S. Li, “Air quality index forecast in beijing based on cnn-lstm multi-model,” *Chemosphere*, vol. 308, p. 136180, 2022.
- [31] Y. Zhou, F.-J. Chang, L.-C. Chang, I.-F. Kao, and Y.-S. Wang, “Explore a deep learning multi-output neural network for regional multi-step-ahead air quality forecasts,” *Journal of cleaner production*, vol. 209, pp. 134–145, 2019.
- [32] S. Chen, “Beijing Multi-Site Air-Quality Data,” UCI Machine Learning Repository, 2019, DOI: <https://doi.org/10.24432/C5RK5G>.
- [33] K. Niwa, H. Chiba, N. Harada, G. Zhang, and W. B. Kleijn, “Microphone array wiener post filtering using monotone operator splitting,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 2036–2046, 2020.
- [34] W. Humphrey, A. Dalke, and K. Schulten, “Vmd: visual molecular dynamics,” *Journal of molecular graphics*, vol. 14, no. 1, pp. 33–38, 1996.
- [35] H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, and W. Zhang, “Informer: Beyond efficient transformer for long sequence time-series forecasting,” in *Proceedings of the AAAI conference on artificial intelligence*, vol. 35, no. 12, 2021, pp. 11 106–11 115.



**Jijing Cai** was born in Zhejiang China in 1998. He is currently pursuing his master's degree at Hangzhou Dianzi University, and jointly trained at Zhejiang A&F University. His research interests include the Internet of Things and Deep learning.



**Tongcun Liu** obtained his Ph.D. in computer science from Beijing University of Posts and Telecommunications in 2019. Currently, he serves as a lecturer at the School of Mathematics and Computer Science, Zhejiang A&F University. His research focuses on data mining, recommender systems, user behavior modeling, and predictive analytics.



**Tingting Wang** received the B.S. degree in Plant Science from Liaoning Normal University, Dalian, China in 2015. She is currently a Ph.D. candidate at Macau University of Science and Technology. Her research interests include the Internet of Things, machine learning, and computational social science.



**Hailin Feng** received his Ph.D. in computer science from the University of Science and Technology of China in June 2007. Since 2007, he worked in the School of Information Engineering of Zhejiang A&F University. From 2013 to 2014, he was a visiting professor at Forest Products Laboratory, USDA. He is currently a professor in the School of Mathematics and Computer Science. His main areas of interest include computer vision, intelligent information processing, and the Internet of Things.



**Kai Fang** (Member, IEEE) received PH.D. from Macau University of Science and Technology in 2023. He is currently a full professor at the School of Mathematics and Computer Science, Zhejiang A&F University. His research interests include IoT, IoV, Computational Social Science, and Deep Learning.



**Ali Kashif Bashir** (Senior Member, IEEE) received the Ph.D. degree in computer science and engineering from Korea University, Seoul, South Korea, in 2012. He is a Leader in the Department of Computing and Mathematics, Manchester Metropolitan University, Manchester, U.K., an Adjunct Professor with Woxsen School of Business, Woxsen University, Hyderabad, India, and also with the Department of Computer Science and Mathematics, Lebanese American University, Beirut, Lebanon. His research interests include the Internet of Things, wireless networks, distributed systems, network/cyber security, network function virtualization, and machine learning. Dr. Bashir is a member of IEEE Industrial Electronic Society, a member of ACM, and a Distinguished Speaker of ACM. He is the Editor-in-Chief of IEEE FUTURE DIRECTIONS NEWSLETTER. He is also an Area Editor of KSII Transactions on Internet and Information Systems and an Associate Editor of IEEE Internet of Things Magazine, IEEE ACCESS, Peer J Computer Science, IET Quantum Computing, and Journal of Plant Diseases and Protection.



**Wei Wang** (Member, IEEE) received the Ph.D. degree in software engineering from the Dalian University of Technology in 2018. He is currently a Professor with Guangdong-Hong Kong-Macao Joint Laboratory for Emotion Intelligence and Pervasive Computing, Artificial Intelligence Research Institute, Shenzhen MSU-BIT University, and also with the School of Medical Technology, Beijing Institute of Technology, Beijing, China. He had been the UM Macao Research Fellow at the University of Macau, Macao SAR. His research interests include computational social science, data mining, the Internet of Things, and artificial intelligence.