AIoT-driven multi-source sensor emission monitoring and forecasting using multi-source sensor integration with reduced noise series decomposition

Mughair Aslam Bhatti[1,2], Zhiyao Song[1,2*], Uzair Aslam Bhatti[3] and Syam M. S[4]

Abstract

The integration of multi-source sensors based AIoT (Artificial Intelligence of Things) technologies into air quality measurement and forecasting is becoming increasingly critical in the fields of sustainable and smart environmental design, urban development, and pollution control. This study focuses on enhancing the prediction of emission, with a special emphasis on pollutants, utilizing advanced deep learning (DL) techniques. Recurrent neural networks (RNNs) and long short-term memory (LSTM) neural networks have shown promise in predicting air quality trends in time series data. However, challenges persist due to the unpredictability of air quality data and the scarcity of long-term historical data for training. To address these challenges, this study introduces the AIoT-enhanced EEMD-CEEMDAN-GCN model. This innovative approach involves decomposing the input signal using EEMD (Ensemble Empirical Mode Decomposition) and CEEMDAN (Complete Ensemble Empirical Mode Decomposition with Adaptive Noise) to extract intrinsic mode functions. These functions are then processed through a GCN (Graph Convolutional Network) model, enabling precise prediction of air quality trends. The model's effectiveness is validated using air pollution datasets from four provinces in China, demonstrating its superiority over various deep learning models (GCN, EMD-GCN) and series decomposition models (EEMD-GCN, CEEMDAN-GCN). It achieves higher accuracy and better data fitting, outperforming other models in key metrics such as MAE (Mean Absolute Error), MSE (Mean Squared Error), MAPE (Mean Absolute Percentage Error), and R^2 (Coefficient of Determination). The implementation of this AIoT-enhanced model in air pollution prediction allows decision-makers to more accurately anticipate changes in air quality, particularly concerning carbon emissions. This facilitates more effective planning of mitigation measures, improvement of public health, and optimization of resource allocation. Moreover, the model adeptly addresses the complexities of air quality data, contributing significantly to enhanced monitoring and management strategies in the context of sustainable urban development and environmental conservation.

Keywords Multi-source forecasting, AIoT sensors, CEEMDAN, GCN

*Correspondence:
Zhiyao Song
zhiyaosong@163.com
Full list of author information is available at the end of the article

Bhatti *et al. Journal of Cloud Computing*   (2024) 13:65

Page 2 of 18

Introduction

Air pollution, a critical issue stemming from economic growth and industrial development, leads to the release of harmful gases and particles into the air. These pollutants are key contributors to smog, acid rain, and the greenhouse effect, which in turn cause adverse weather conditions, global temperature changes, and detrimental effects on ecosystems [1]. Furthermore, air pollution poses substantial health risks, including respiratory disorders, lung cancer, and exacerbation of conditions like asthma and allergies [2]. Urbanization amplifies these challenges in densely populated areas.

The role of AIoT (Artificial Intelligence of Things) has become increasingly significant in understanding and managing air pollution, particularly in monitoring and analyzing carbon emissions. AIoT technologies enable more sophisticated and comprehensive monitoring of air quality by integrating automated monitoring stations with advanced data analytics [3]. This integration facilitates real-time tracking of pollutant emissions, offering granular insights into their sources and dispersion patterns. Air quality forecasting, crucial for effective pollution control, has evolved with the application of AIoT. This approach enhances traditional statistical models, including linear and nonlinear machine capturing algorithms [4], with deep learning techniques. These AI-driven models excel in extracting relevant features and capturing temporal dependencies in air quality data [5, 6].The explosion of time series data, fueled by technological advancements, industrialization, and the proliferation of sensors, has been a game changer in this field. Time series data, a chronological sequence of data points, is instrumental in various aspects of life, including environmental monitoring [7, 8]. AIoT technologies significantly enhance the utility of time series data by enabling more efficient processing and analysis, leading to better predictions and understanding of air quality trends.

Advanced forecasting techniques, augmented by AIoT, provide critical insights for researchers and policymakers. These insights are essential for designing effective pollution management and mitigation strategies, improving the understanding of the complex interplay between human activities, atmospheric conditions, and the environment. The proposed EEMD-CEEMDAN-GCN model, enhanced by AIoT, addresses the challenges in accurately capturing the dynamics of air quality data. It represents a breakthrough in the field of air quality forecasting, offering a sophisticated approach to dealing with the intricacies of atmospheric data, with a particular focus on carbon emission monitoring. This model's integration with AIoT technologies marks a significant advancement in emissions monitoring, setting a new standard for accuracy and efficiency in environmental management and sustainability efforts.

The EEMD-CEEMDAN-GCN model, an innovative hybrid predictive model for time-varying forecasting, is introduced in the paper. To increase the precision and efficacy of time series predictions, this model combines the principles of signal breakdown and deep learning. The EEMD-CEEMDAN-GCN hybrid prediction model combines the Graph Convolutional Network (GCN), a deep learning method, with the Ensemble Empirical Mode Decomposition (EEMD) and Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) signal decomposition approaches. The EEMD and CEEMDAN algorithms are used to break down the input time series data to its inner mode functions (IMFs) and excess noise. These IMFs and residual noise are then used to forecast using the GCN network. The application of signal decomposition techniques supports extracting meaningful information from time series data and improves prediction accuracy. The suggested model has potential applications in a variety of industries, like finance, weather forecasting, and energy demand prediction, where accurate time series prediction is crucial for decision-making. It can also be used for anomaly detection and fault diagnosis in industrial processes. We may extract more valuable information from the data and increase the precision of our predictions by dissecting the input signal into several components using a variety of methods. Additionally, by simplifying the original time series data, this method can make it simpler to analyze and interpret. Overall, employing a number of signal decomposition techniques is a potent method for time series analysis with applications in a wide range of disciplines, including engineering, environmental science, and finance. The study's primary contributions are:

- Two advanced signal decomposition algorithms, Ensemble Empirical Mode Decomposition (EEMD) and Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN), are employed within an AIoT framework to dissect different components of the same time series information, specifically focusing on emissions data. This process significantly refines the data representation, uncovering underlying patterns and simplifying the complexity of the time series related to emissions.
- The innovative EEMD-CEEMDAN-GCN model, augmented with AIoT capabilities, is then utilized to predict target variables, such as emission levels,

Bhatti *et al. Journal of Cloud Computing*   (2024) 13:65

Page 3 of 18

based on the decomposed components. This model integrates the Graph Convolutional Network (GCN) deep learning algorithm, leveraging its strength in learning and capturing temporal dependencies within emission data, thereby enabling more accurate and reliable prediction.

- To evaluate the effectiveness of this AIoT-enhanced model, experimentation is conducted using real-world datasets, particularly the Air Quality dataset, which is critical in studying emissions. The performance of the EEMD-CEEMDAN-GCN model is benchmarked against other prevalent algorithms, including GCN, EMD-GCN, EEMD-GCN, CEEMDAN-GCN, and EMD-CEEMDAN-GCN, providing a comprehensive comparison in the context of emissions time series prediction.

The rest of the paper is organized as follows: Literature review section focuses on the latest literature about emissions and predictions. Proposed methodology section focuses on the proposed method. EEMD Algorithm section discusses results while CEEMDAN Algorithm section provides a conclusion and future implications.

Literature review

Time series data prediction has become a crucial aspect of extensive data analysis, profoundly impacting the transformation and optimization of various living and industrial activities, especially in the context of AIoT (Artificial Intelligence of Things) and emissions management [9, 10]. The integration of AIoT in these analyses significantly enhances the ability to monitor and predict environmental factors, particularly emissions, with greater accuracy and efficiency.

In the realm of emissions forecasting, Patra [11] utilized multi-layer perceptron (MLP), support vector regression (SVR), and autoregressive integrated moving average (ARIMA) algorithms for one-month carbon dioxide and nitrogen dioxide predictions. This study used the public Air Quality database from the UCI Machine Learning Repository [12], combined with data from 5 metal oxide chemical detectors in an Air Quality Chemical Multisensory Device, producing 390 instances of daily mean responses. The AIoT framework plays a crucial role here, enabling the integration and analysis of diverse data sources for more precise emissions predictions.

Bekkar et al. [13] proposed a GCN-based artificial intelligence framework, leveraging AIoT in smart city contexts for enhanced air pollution forecasting. This approach utilizes IoT data to provide real-time, accurate predictions of air quality, focusing on emissions monitoring and control.Feng H et al. [14] introduced an encoder-decoder model based on deep learning to address data gaps in air quality and meteorological series, with a focus on South Korea. This model, enhanced by AIoT capabilities, significantly improves the prediction of air quality, particularly emissions data, by handling missing values more effectively.

Waseem K H et al. [15] compared the performance of RNN, GCN, and gated recurrent unit (GRU) [16] networks for air pollution prediction using AirNet data, with an emphasis on emissions. The integration of these models with AIoT technologies facilitates more sophisticated analysis and forecasting of pollution levels. Qi Z et al. [17] developed a deep air adaptable technique that combines feature selection with a spatiotemporal semi-supervised neural network, an approach that can greatly benefit from AIoT in capturing the dynamic nature of emissions data across different geographies. Ahmed M et al. [18] estimated the Air Quality Index (AQI) using an RNN-GCN model, an approach that can be significantly enhanced by incorporating AIoT for real-time emissions monitoring and prediction. Lastly, Masih [19] reviewed machine learning techniques in environmental science and engineering research, highlighting the importance of ensemble learning, linear regression, neural networks, and SVM for pollution estimation and forecasting tasks. The integration of these techniques with AIoT technologies presents a significant advancement in emissions monitoring, offering more accurate, efficient, and timely predictions, which is crucial for effective environmental management and policy-making.

The studies evaluate the efficiency of the models using metrics that is RMSE, MAE, and R^2, demonstrating the effectiveness of deep learning approaches in air pollution estimation and forecasting.

According to the relevant literature, developing a numerical model poses challenges due to meteorological systems' complex and uncertain nature, resulting in low forecast accuracy. The ability of statistical models to predict unpredictable regressive sequence data depends on the data's consistency. When approximating nonlinear sequential data, machine-learning and deep-learning techniques offer adaptive skills and advantages. However, they still need help in learning and achieving high prediction accuracy with nonstationary data. Furthermore, when neural networks are employed directly for modeling the Atmospheric Quality measure (AQI), which is a composite measure that inherits the fluctuation and variability characteristics of the meteorological framework,

Bhatti *et al. Journal of Cloud Computing*   (2024) 13:65

Page 4 of 18

it has a detrimental impact on prediction models and results in low accuracy.

Researchers have explored integrated prediction models to overcome these limitations to enhance forecast stability and accuracy. The empirical modal decomposition approach, specifically complementary ensemble empirical mode decomposition-SVR (CEEMD-SVR) [20], has shown promising results in predicting $PM_{2.5}$ mass concentration. The CEEMD-Elman model [21] rely on empirical mode decomposition (EMD) has provided a foundation for successful AQI trend prediction. Techniques involving complementary sets [22] EMD combined with GCN neural networks have been proposed for enhancing short-term power load prediction accuracy. A air velocity combined estimation process [23] based on empirical ensemble mode decomposition (EEMD) has been developed to enhance EMD mode mixing. Additionally, combining GCN model with signal decomposition techniques, such as EMD, has proven effective in improving prediction accuracy for various applications, including hourly concentration prediction [24].

The study underscores the importance of meticulously analyzing and preprocessing the original data before developing predictive models, a process significantly enhanced by AIoT (Artificial Intelligence of Things) technologies. EEMD (Ensemble Empirical Mode Decomposition) and its advanced iteration, CEEMDAN (Complete Ensemble Empirical Mode Decomposition with Adaptive Noise) [25, 26], are identified as algorithms adept at tackling these preprocessing challenges. When combined with AIoT, these methods can handle large-scale data from various IoT sensors, ensuring more refined and accurate data preparation for subsequent analysis. Hybrid models that merge EEMD or CEEMDAN with deep learning techniques have demonstrated increased accuracy in applications like financial time series forecasting and short-term stock price trend prediction [26]. The integration of these models with AIoT platforms enables the handling of vast and complex data sets, typical in financial markets, enhancing prediction accuracy and reliability.

GCN (Graph Convolutional Network) is recognized as an effective strategy for predicting chaotic time series [27–29], and its application within an AIoT framework allows for more sophisticated analysis of data characterized by high volatility and unpredictability. This is particularly useful in industries where data is influenced by a multitude of interconnected factors, such as energy or traffic management. The "decomposition before reconstruction" paradigm, primarily utilizing

EEMD and CEEMDAN [30], has proven successful in various forecasting domains, including PM2.5 prediction and long-term stream-flow forecasting. When these decomposition methods are applied in conjunction with AIoT, they bring additional benefits such as the ability to process large volumes of environmental data, effectively overcoming mode mixing issues and achieving low reconstruction errors. This makes them particularly suitable for time series decomposition in studies where IoT devices are used for environmental monitoring and data collection. The integration of these methods with AIoT technologies thus enhances the overall efficiency and accuracy of time series analysis, especially in applications requiring high precision, such as environmental monitoring and resource management.

Additionally, a number of investigators have recently highlighted to the benefits of a graph neural network in fields like flow of traffic estimation [31–34], parking availability prediction [35], pedestrian trajectories prediction [36, 37]. These benefits have also been applied to other domains, such as air quality, and several authors have employed neural networks with graphs for predicting air quality. These benefits have also been applied to other domains, such as air quality, and a few researchers have employed graph neural networks for forecasting air quality. Using records for the Beijing-Tianjin-Hebei and Pearl River Delta urban areas, Han et al. [38] put forward the Self-Supervised a hierarchy Graph Neural Network based on cities-functional zones-regions hierarchical graph network to perform extremely fine air quality prediction. To perform the Air Quality Index (AQI) predictions, Ram et al. [39] proposed a Dual GCN (DGCN) and LSTM network combined with a wireless sensor network and Internet of Things (IoT). DGCN aids in processing the sensor's data that was later processed by the graph LSTM [40–43].

The existing literature on air quality prediction models reveals a substantial research gap in effectively addressing the unpredictable and nonlinear nature of atmospheric conditions, leading to challenges in achieving accurate predictions. While statistical, machine learning, and deep learning models have been extensively employed, limitations persist. To bridge this gap, a novel approach integrating empirical mode decomposition (EMD) techniques, specifically complementary ensemble empirical mode decomposition (CEEMDAN) and enhanced empirical mode decomposition (EEMD), with the Graph Convolutional Network (GCN) is introduced. GCN is used to effectively record the topological data of the whole monitoring network. This innovative hybrid model,

EEMD-CEEMDAN-GCN, aims to overcome the short-comings of traditional models and enhance prediction accuracy in the dynamic field of air pollution forecasting. The proposed model's application to real-world datasets, specifically the Air Quality dataset, substantiates its superiority over existing methods like GCN, EMD-GCN, EEMD-GCN, CEEMDAN-GCN, EMD-CEEMDAN-GCN and EEMD-CEEMDAN-GCN.

Proposed methodology

Before describing the EEMD-CEEMDAN-LSTM hybrid model used for time series data prediction, we will briefly discuss the related fundamental theories behind this model construction, namely, the Ensemble Empirical Mode Decomposition (EEMD), the Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN), and the principles and applications of the Long Short-Term Memory Network (LSTM). The methodology employed in this study revolves around the innovative integration of the Ensemble Empirical Mode Decomposition (EEMD), Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN), and Graph Convolutional Network (GCN) components to form the EEMD-CEEMDAN-GCN model. This section outlines the rationale behind selecting this novel model, elucidates the step-by-step development process, and delineates the distinctive roles played by each constituent element.

EEMD Algorithm

EEMD is an improved method based on Empirical Mode Decomposition (EMD). EEMD adds Gaussian white noise to the original data for EMD decomposition, which successfully solves the problems of mode mixing and endpoint effects in traditional EMD. The first stage involves applying EEMD to the input signals. EEMD decomposes signals into IMFs by iteratively adding noise to the original signal and extracting the resulting oscillatory components. This process ensures adaptability to diverse signal characteristics and enhances the model's robustness in capturing complex patterns. The algorithm framework of EEMD is shown in Algorithm 1.The calculation process of the EEMD algorithm is as follows:

1) Add Gaussian white noise to the original data to generate a new data set, as shown in formula (1).

$$x_i(t) = x(t) + n_i(t), i = 1, 2, L, M \tag{1}$$

where $n_i(t)$ is the white noise data added at the i-th time, x(t) is the original data, and $x_i(t)$ is the new data with added white noise generated at the i-th time.

2) Decompose the new sequence data with added white noise, $x_i(t)$, into n Intrinsic Mode Functions (IMFs) and a residual component using the EMD method, as shown in formula (2):

$$x_i(t) = \sum_{j=1}^{n} f_{i,j}(t) + r_{in}(t) \tag{2}$$

Here, $\sum_{j=1}^{n} f_{i,j}(t)$ is the j-th IMF obtained after adding white noise for the i-th time, while $r_{in}(t)$ represents the residual component obtained after adding white noise, which represents the average trend of the signal. n represents the number of IMF components.

Begin:

Setting the parameters for the number of operations m and adding white noise $n_i(t)$
    For i = 1; i ≤ m; i++;
    Add $n_i(t)$ to the signal $x(t)$: $x_i(t) = x(t) + n_i(t)$;
    Getting the modal components and corresponding residual components of $x_i(t)$ through EMD decomposition;
    End For

Taking the average and accumulating all the IMF components and residual components.
End

---

3) Repeat the above steps m times, adding different white noise each time, to obtain n decomposition results of m sequences:

$$\sum_{i=1}^{m} \left[\sum_{j=1}^{n} (t) + r_{in}(t) \right] \tag{3}$$

4) Utilizing the characteristic of Gaussian white noise having a mean of zero, take the average of each IMF component and residual component obtained from the previous steps, and sum them up to obtain the final result, as shown in formula (4):

$$x(t) = \sum_{j=1}^{n} \left[\frac{1}{m} \sum_{i=1}^{m} f_{ij}(t) \right] + \frac{1}{m} \sum_{i=1}^{m} r_{in} \tag{4}$$

EEMD primarily serves to decompose input signals into a set of IMFs, providing a meaningful representation of the underlying patterns within the data. Its adaptive nature ensures flexibility in handling signals with varying complexities.

CEEMDAN Algorithm

The CEEMDAN algorithm, as shown in Algorithm 2, is also an improved method based on EMD. It overcomes the mode mixing problem in EMD. Unlike EEMD, CEEMDAN does not directly add Gaussian white noise to the original signal but adds auxiliary noise to the mode components obtained after EMD decomposition. At the same time, the overall average calculation begins after obtaining the first mode component and continues until obtaining the final mode component. This process is then repeated for the residual component. CEEMDAN builds upon EEMD by introducing an adaptive noise correction mechanism. It refines the extracted IMFs by iteratively reducing the noise content, improving the accuracy of signal decomposition. This step is crucial for obtaining a set of clean and representative IMFs that can be effectively utilized in subsequent analysis.

The calculation process of the CEEMDAN algorithm is as follows:

1) Add Gaussian white noise to the original signal xt, as shown in formula (5):

$$x^i(t) = x(t) + \varepsilon_0 n^i(t), i = 1, 2, \cdots, N \qquad (5)$$

where ε_0 represents signal-to-noise ratio, $n^i(t)$ represents the Gaussian white noise added at the i-th time, and N represents the total number of experiments.

2) Perform EMD decomposition on each new signal $x^i(t)$ with added Gaussian white noise to obtain the first Intrinsic Mode Function IMF_1 and the residual component r_1, as shown in formulas (6) and (7):

$$IMF_1 = \frac{1}{N} \sum_{i=1}^{N} E_1\left[f^i(t)\right] \qquad (6)$$

$$r_1(t) = x(t) - IMF_1 \qquad (7)$$

Here, E represents the EEMD decomposition operation.

CEEMDAN refines the output of EEMD by mitigating noise and enhancing the accuracy of signal decomposition. This step is essential for improving the fidelity of the extracted IMFs, making them suitable for subsequent analysis.

Begin:

Setting the number of experiments N and the parameters of adding white noise $n_i(t)$.

Add $n_i(t)$ to the signal $x(t)$: $x_i(t) = x(t) + n_i(t)$;
For i=1; i ≤ N; i++;

  Getting the i-th IMF by EEMD;
   Add Gaussian white noise to i-th IMF;
  Calculating the i-th residual component through the i-th IMF;
  Add Gaussian white noise to i-th residual component;
  If the signal cannot decomposed by EMD:

   Break;

End For

Obtain all IMF and Residue.

End

3) Perform EMD decomposition on $r_1(t)$ with added $\varepsilon_1 E_1[n^i(t)]$, and obtain the IMF_2, as shown in formula (8):

$$IMF_2 = \frac{1}{N} \sum_{i=1}^{N} E_1\left\{r_1(t) + \varepsilon_1 E_1[n^i(t)]\right\} \qquad (8)$$

4) When $k = 2, 3, \cdots, K$, calculate the k-th residual component $r_k(t)$, as shown in formula (9):

$$r_k(t) = r_{k-1}(t) - IMF_k \qquad (9)$$

5) Add white noise to form a new time series in each stage and calculate the first intrinsic mode function of this time series as the new mode component of the original time series. Then, the k-th stage mode component IMF_{k+1} is calculated, as shown in formula (10):

$$IMF_{k+1} = \frac{1}{N} \sum_{i=1}^{N} E_1\left\{r_k(t) + \varepsilon_k E_k[n^i(t)]\right\} \quad (10)$$

6) Repeat steps 4 and 5 to ensure that the signal cannot be further decomposed by EMD and obtain k-mode components. The final residual component of the signal is:

Bhatti *et al. Journal of Cloud Computing*   (2024) 13:65

Page 7 of 18

$$R(t) = x(t) - \sum_{k=1}^{K} IMF_k \tag{11}$$

7) The signal $x(t)$ can be represented by CEEMDAN decomposition as follows:

$$x(t) = \sum_{i=k}^{K} IMF_{k+1} + R(t) \tag{12}$$

GCN neural network

Defining a graph: $G = (V, E)$, where V is the set of all nodes in the graph, and E is the set of all edges between the nodes in the graph. The essence of a graph convolutional network is to extract data features of graph-type data. Unlike convolutional neural networks, graph convolutional networks can capture spatial characteristics of topological graphs in non-Euclidean spaces and extract the correlation between data in the graph structure. The core idea of the graph convolutional network is to construct a message-passing mechanism similar to that of the graph convolutional network. From the original graph structure data, by continuously extracting features and passing data, the data information of the target node itself and its neighborhood space are continuously updated, Following the signal decomposition, the IMFs are transformed into a graph structure, where nodes represent individual components, and edges capture relationships between them. The GCN is then applied to learn and exploit the inherent graph topology, enabling the model to capture long-range dependencies and intricate relationships among the extracted IMFs.

The state update formula of GCN is as follows:

$$D_{ii} = \sum_j A_{ij} \tag{13}$$

$$H^{(0)} = X \tag{14}$$

$$H^{(1)} = \sigma(D^{-1/2}(A + I)D^{-1/2}XW^{(0)}) \tag{15}$$

$$H^{(l+1)} = \sigma(D^{-1/2}(A + I)D^{-1/2}H^{(l)}W^{(l)}) \tag{16}$$

In the formula, A represents the adjacency matrix that describes the spatial relationship of the original data; I is the identity matrix; $A + I$ represents an adjacency matrix with self-connections added; D is the degree matrix, which can represent the relationship between each node and other nodes in the graph. For a node, the more nodes it is associated with, the higher the corresponding value in the degree matrix. The term $D^{-1/2}(A + I)D^{-1/2}$ represents the normalization operation of the adjacency matrix. Its role is to improve the problem of gradient vanishing or explosion that often occurs in deep learning. X is a feature matrix composed of vertex feature data, and W represents the connection weight of each edge; $H^{(1)}$ represents the result obtained from the first message passing, and similarly, $H^{(l+1)}$ represents the result obtained from the $l+1$-th update. $W^{(l)}$ is the connection weight parameter obtained by aggregating and updating the parameters. The initialization of the W parameter in the GCN algorithm is not strictly required. Compared with other deep learning models, GCN can achieve the purpose of effectively updating node features by stacking shallow layers and has lower parameter volume and computation time complexity. The GCN component harnesses the power of graph-based learning, leveraging the relationships between the decomposed IMFs. By capturing dependencies and non-linear interactions, GCN enhances the model's ability to discern intricate patterns and features within the signal data.

Proposed EEMD-CEEMDAN-GCN model

The model proposed in this paper (Fig. 1) achieves high-precision prediction of time series data by combining the EEMD and CEEMDAN signal decomposition methods with the GCN deep learning method. Time series data has complex characteristics such as dynamic non-linearity and non-stationary, among others. If a single prediction model is used, it cannot achieve ideal results. EEMD and CEEMDAN signal decomposition techniques can decompose complex time series signals into multiple modal components with lower complexity. Therefore, adding signal decomposition steps to the time series prediction method improves prediction accuracy. Previous studies often used only one decomposition method in the signal decomposition process, which cannot entirely reduce the complexity of the signal and learn the correlation between different components under the same data. The choice of the EEMD-CEEMDAN-GCN model stems from the need for a comprehensive approach that combines signal decomposition, noise reduction, and graph-based learning. EEMD and CEEMDAN are powerful techniques for decomposing complex signals into intrinsic mode functions (IMFs), enabling the extraction of inherent patterns. The subsequent integration of GCN facilitates the exploitation of relationships within the data, enhancing the model's ability to capture intricate dependencies and non-linearities. The EEMD-CEEMDAN-GCN model combines the strengths of signal decomposition, noise reduction, and graph-based learning to provide a comprehensive framework for extracting and understanding complex patterns within data. This methodology ensures a robust and adaptive approach, making it a significant contribution to the field.
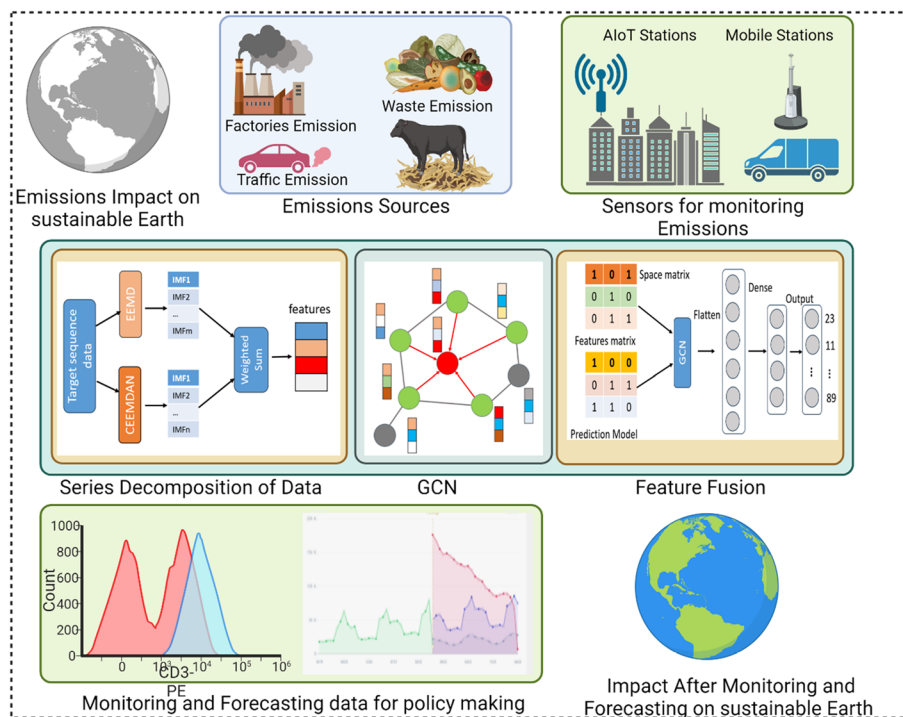
Fig. 1 The flow chart of the proposed model

In this study, the EEMD and CEEMDAN signal decomposition methods were combined in a supplementary way to fully decompose the input data and screen the modal components generated by the signal decomposition using the correlation between the target sequence data to be predicted and other component sequence data. Finally, combined with the excellent performance of the GCN neural network in representing time series autocorrelation and long-term memory, the model can achieve superior prediction accuracy.

The detailed modeling steps are as follows:

1) Preprocess the raw data by filling missing values with the mean and ordering all data in a unified standard time sequence.
2) Decompose each component of the time series data by EEMD to obtain their respective modal components. Calculate the correlation coefficient between these modal components and the target time series data to be predicted and select the modal components with a correlation coefficient greater than 0.35.
3) Decompose the target time series data to be predicted by CEEMDAN to obtain its corresponding series of modal components and residual components.
4) Input the modal components obtained from steps 2 and 3 as features into the GCN deep neural network

to obtain the final prediction result. The flow chart of the proposed model is shown in Fig. 1.

Data sets & evaluation metrics
Study area
Anhui, Guangdong, Henan, Hubei, and Jiangsu are provinces in China with diverse geographical features and unique characteristics. Anhui is a hilly and mountainous province known for its lakes and waterways, while Guangdong experiences a subtropical climate and has a varied economic growth pattern. Henan is located in the middle and lower reaches of the Yellow River, known for its agricultural productivity. Hubei gained global attention as the epicenter of the COVID-19 outbreak and has a transitional topography. Jiangsu, an economically developed province, has faced air quality challenges due to rapid economic expansion (Fig. 2).

Data sets
The air pollution index is based on ambient air quality criteria and the impact of different contaminants on human health, ecology, and the environment. The concentration of multiple air contaminants that are consistently detected is reduced to a single conceptual index value. Based on research from the past, two primary air pollutants ($PM_{2.5}$, PM_{10}) are closely connected with air pollution.

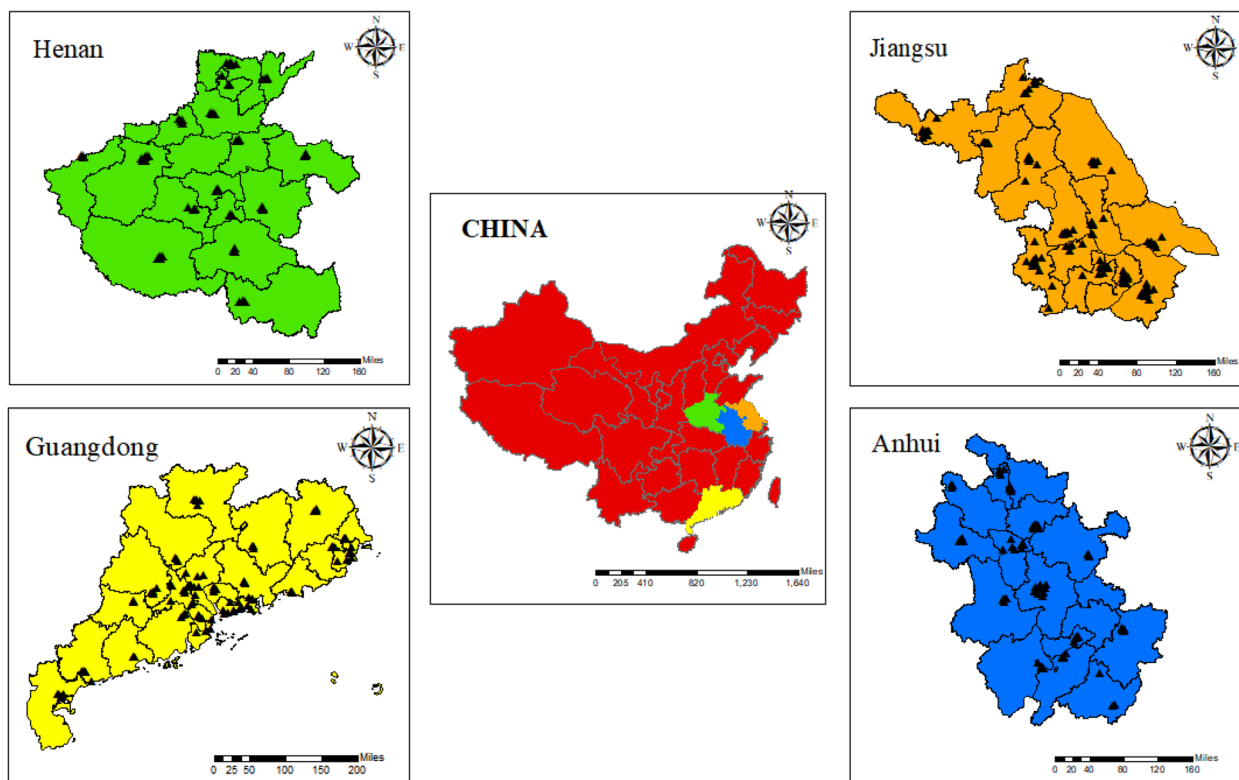Bhatti *et al. Journal of Cloud Computing*   (2024) 13:65

Page 9 of 18

Fig. 2 The study area map of Henan, Jiangsu, Guangdong, Anhui

To prove the validity and robustness of the hybrid model, this study investigated 68 cities from Anhui (14 cities), Guangdong (21 cities), Henan (16 cities), Hubei (4 cities), and Jiangsu (13 cities) with the worst air quality in China based on primary pollutants, economic conditions, and geographical factors. The datasets utilized in this study are all open-source data taken from the China Environmental Monitoring Station's national urban air quality real-time release platform (http://www.cnemc. cn/). Between January 1, 2017, and August 14, 2021, air quality data were collected daily at monitoring locations in Anhui, Guangdong, Henan, Hubei, and Jiangsu. A statistical description of the data for each province is shown in supplementary material A, and statistical graphs of data description in supplementary materials B.

Correlation between different air quality pollutants of each province is shown in supplementary materials C to show the relationship and variation of pollutants.

The study encompasses 68 cities distributed across the selected provinces, with the distribution as follows: Anhui (14 cities), Guangdong (21 cities), Henan (16 cities), Hubei (4 cities), and Jiangsu (13 cities). The selection criteria for these cities were based on the severity of air pollution, primary pollutants, economic

conditions, and geographical factors, ensuring a diverse and representative dataset.

The datasets are characterized by their substantial size, capturing daily air quality measurements over a four-year period. The inclusion of multiple provinces and cities ensures diversity, encompassing various geographical and economic conditions. The representativeness of the dataset is underlined by the deliberate selection of cities with the worst air quality in China, providing a comprehensive view of challenging environmental conditions.

Prior to model training, the datasets were divided into training (80%), validation (10%), and testing (10%) sets in an 8:1:1 ratio. This partitioning scheme ensures an adequate amount of data for model training while maintaining distinct sets for validation and testing to assess the model's generalization capabilities.

Figure 2 presents a study area map, illustrating the geographical distribution of the selected provinces. Supplementary materials A provide statistical descriptions of the data for each province, while supplementary materials B include statistical graphs for a detailed data description. Supplementary materials C showcase the correlation between different air quality pollutants

Bhatti *et al. Journal of Cloud Computing*   (2024) 13:65

Page 10 of 18

in each province, elucidating the relationships and variations among pollutants.

Table 1 delineates the parameter values for various models employed in the study, including LSTM, Bi-LSTM, VMD-LSTM, EEMD-LSTM, and EEMD-CEEMDAN-LSTM. The parameters include window length, batch size, dropout rate, learning rate, and the number of epochs. These settings were carefully chosen to optimize the performance of each model and ensure fair comparisons between them.

Evaluation metrics

In this paper, four evaluation metrics were selected to evaluate the prediction performance of the proposed models, namely: Mean Absolute Error (MAE), Mean Square Error (MSE), Mean Absolute Percentage Error (MAPE), and R^2 (R Squared). Their formulas are as follows:

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \widehat{y_i}| \tag{17}$$

$$MSE = \frac{1}{n \sum_{i=1}^{n} (y_i - \widehat{y_i})^2} \tag{18}$$

$$MAPE = \frac{\frac{1}{n} \sum_{i=1}^{n} |y_i - \widehat{y_i}|}{y_i} \tag{19}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \widehat{y_i})^2}{\sum_{i=1}^{n} (y_i - \bar{y})} \tag{20}$$

y_i signifies the actual value of the time series sample, $\widehat{y_i}$ denotes the model's predicted value, n means the number of testing samples, and i represents the sequence number of the testing sample in the above formulae.

Table 1 Parameters setting for the algorithms used in this study

Parameters	LSTM	Bi-LSTM	VMD-LSTM	EEMD-LSTM	EEMD-CEEMDAN-LSTM
Number of IMFs				6	6
Alpha			2000		
Win length	3	3	3	3	3
Batch size	64	64	64	64	64
Drop out	0.2	0.2	0.2	0.2	0.2
Learning rate	0.008	0.008	0.008	0.008	0.008
Epochs	500	500	500	500	500

Results and discussion

We conducted experiments for Air Quality on datasets of Anhui (14 cities), Guangdong (21 cities), Henan (16 cities), and Jiangsu (13 cities) Provinces and compared the performance of GCN, EMD-GCN, EEMD-GCN, CEEMDAN-GCN, EMD-CEEMDAN-GCN and the proposed EEMD-CEEMDAN-GCN in this paper. The performance of these five models on the particulate matter datasets is shown in Figs. 3 and 4. At the same time, a more detailed analysis of each province is available in supplementary material D. The results of the experiments demonstrate that EEMD-CEEMDAN-GCN outperforms the other models in terms of predictive accuracy for air quality. By leveraging the strengths of EEMD and CEEMDAN in decomposing the time series and GCN in capturing temporal dependencies, the proposed approach provides a robust framework for air quality analysis and prediction.

Comparison between the predicted and actual data in Figs. 5 and 6 likely evaluate the performance of the proposed model with other prediction models on PM_{10} and $PM_{2.5}$ datasets, respectively, for the last 150 data points. The purpose of such comparisons is to assess the accuracy and effectiveness of each model in capturing the underlying patterns and making accurate predictions.

The performance of the EEMD-CEEMDAN-GCN model is compared with other models using multiple evaluation metrics, including Mean Absolute Error (MAE), Mean Squared Error (MSE), Mean Absolute Percentage Error (MAPE), and R-squared ($R2$) on the Air Quality dataset. The results demonstrate the superiority of the proposed model across various metrics.

In terms of MAE, EEMD-CEEMDAN-GCN outperforms other models as follows:

19.31% better than GCN, 3.75% better than EMD-GCN, 0.16% better than EEMD-GCN, 5.81% better than CEEMDAN-GCN, 13.10% better than EMD-CEEMDAN-GCN.

For the MSE metric, the EEMD-CEEMDAN-GCN model excels:

45.12% better than GCN, 9.36% better than EMD-GCN, 6.94% slightly worse than EEMD-GCN, 0.97% better than CEEMDAN-GCN, 17.33% better than EMD-CEEMDAN-GCN.

The MAPE metric shows the superiority of EEMD-CEEMDAN-GCN:

24.37% better than GCN, 4.20% better than EMD-GCN, 3.10% better than EEMD-GCN, 13.08% better than CEEMDAN-GCN, 20.16% better than EMD-CEEMDAN-GCN.
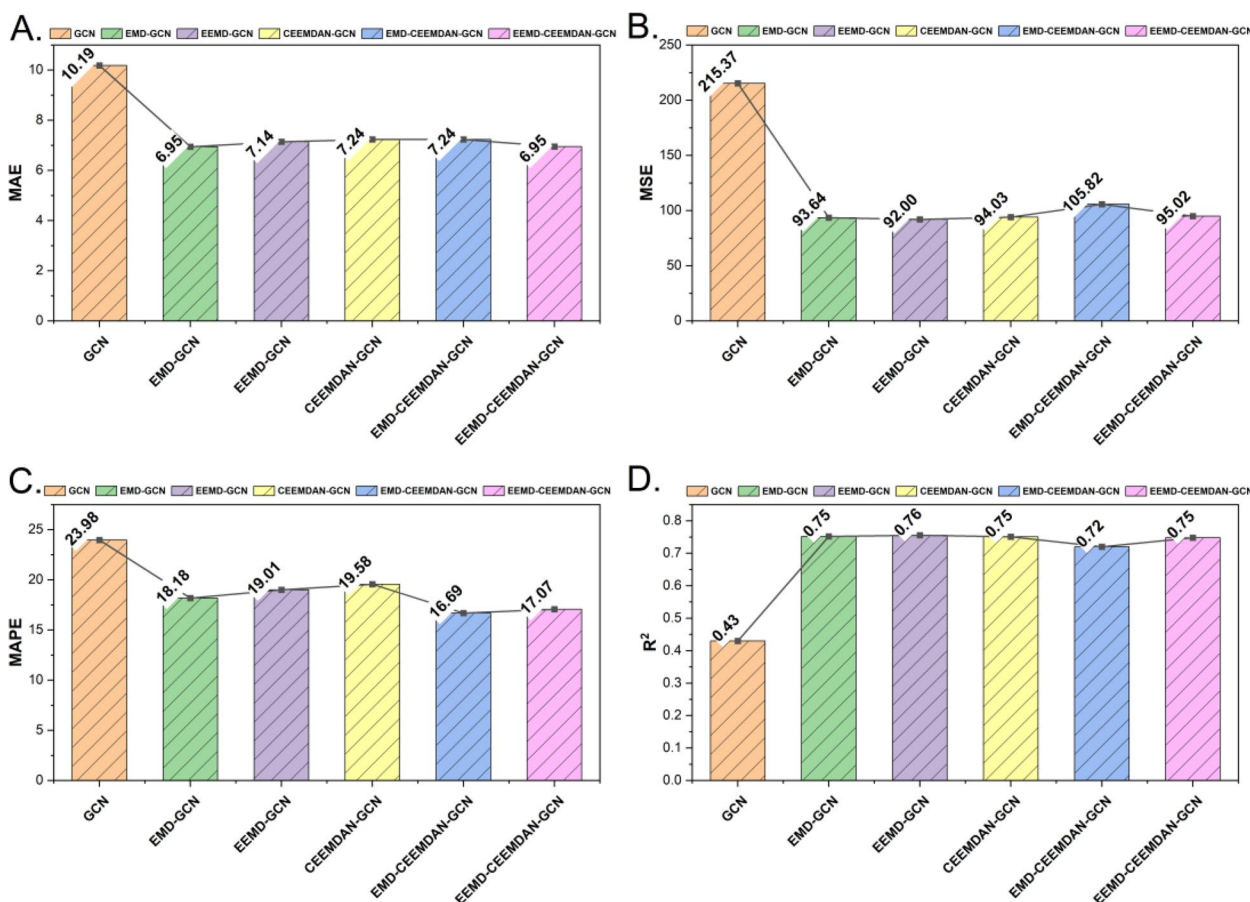
Bhatti *et al. Journal of Cloud Computing* (2024) 13:65

Page 11 of 18

Fig. 3 Evaluation metrics comparison of all the algorithms for PM2.5 datasets remaining are all correct

In terms of R2, EEMD-CEEMDAN-GCN demonstrates superior performance:

> 70.64% better than GCN,—13.95% worse than EMD-GCN, 2.38% slightly equal to EEMD-GCN, 0.33% slightly equal to CEEMDAN-GCN, 15.20% better than EMD-CEEMDAN-GCN.

These results collectively underscore the efficacy of the EEMD-CEEMDAN-GCN model, showcasing its robustness and superior predictive capabilities across multiple evaluation metrics on the Air Quality dataset.

The experiments showed that on the time series air pollution dataset, the proposed EEMD-CEEMDAN-GCN model had the approximately same R^2 as CEEMDAN-GCN and outperformed all other classical prediction models regarding MAE, MSE, and MAPE. The indicated model had performance slightly equal to EEMD-GCN only in terms of MAE, but it significantly outperformed GCN, EMD-GCN, and EEMD-CEEMDAN-GCN in terms of MAE, MSE, MAPE, and R^2. On the Air Quality data set the proposed model outperformed in terms of MAE, MSE, MAPE, and R^2. Figures 7 and 8 show the

spatiotemporal change of R^2 in different cities after making predictions with different methods. In contrast, further detailed performance metric results are shown in supplementary material E.

It has been difficult to decipher the complexities of time series data in air pollution research. To improve the precision of prediction models, a group of forward-thinking researchers investigated decomposition approaches. Due to the dangers of modal aliasing and the inefficiency of approaches like EMD, EEMD, and VMD, previous solutions have failed. Here comes CEEMDAN, a glimmer of light in the realm of air pollution time series decomposition. Its appearance signaled a turning point because it finally addressed the problems that the initial decomposition methodology had long struggled with. In a groundbreaking work, [44] decomposed the IMF in a simplex manner and found that the initial component contained the most complexity.

Inspired by this revelation, our study aimed to quantify the difficulty inherent in each IMF component using the formidable CEEMDAN technique. Through our endeavors, we have established quantitative criteria for
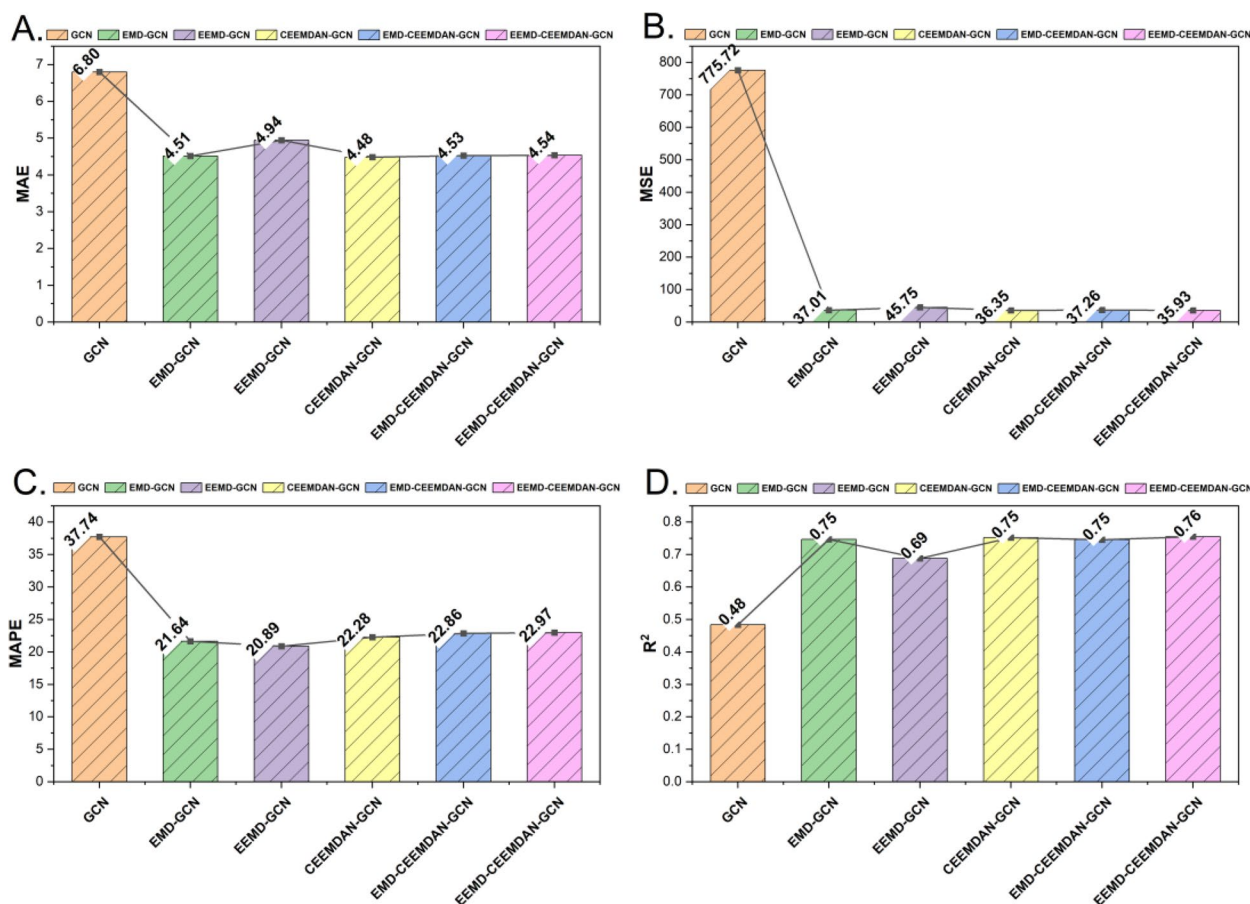
Fig. 4 Evaluation metrics comparison of all the algorithms for PM_{10} datasets

identifying the most intricate segments, paving the way for a more refined selection process.

Building upon the achievements of Li P et al., [45], who employed decomposition-integrated frameworks, our models have witnessed a significant boost in predictive performance. Taking the pollution data from Anyang city as our guide, the EEMD–GCN approach outshone its GCN counterpart, showcasing improvements of 50.8%, 51.81%, and 52.96% in terms of MAE, RMSE, and MAPE, respectively. The EEMD-GCN and CEEMDAN–GCN models also exhibited remarkable prediction performance only in some metricses, early warning accuracy [46], and stability across numerous datasets. Our findings align with this research, further confirming the superiority of EEMD–GCN over GCN after series decomposition.

Another exploration avenue involved removing noise from air quality data, courtesy of Huang et al. [47] and their EMD model. The elusive IMF components were extracted through this process, leading us to fashion an EMD–IPSO–GCN air quality prediction model for each constituent. Validation analyses of this algorithm reinforced its theoretical and technological underpinnings,

showcasing heightened prediction affivacy and showing a better model fitting compared to GCN and EMD–GCN. Our study echoes these sentiments, presenting a comparative approach using CEEMDAN, outperforming GCN and EMD–GCN.

The fusion of EEMD or CEEMDAN with deep learning methodologies in hybrid models has yielded substantial advancements in financial time series forecasting and short-term stock price trend prediction. GCN, renowned for its prowess in predicting the chaotic time series, is a fitting partner in constructing a chaotic time series prediction methods. The "decomposition before reconstruction" methodology has shown success in a number of forecasting areas, including $PM_{2.5}$ forecasting and long-term streamflow forecasting, in part through utilizing EEMD and CEEMDAN. These decomposition techniques offer unparalleled benefits, triumphing over mode mixing and yielding low reconstruction errors, positioning them as apt choices for time series decomposition in our investigation.

The consensus among academics holds that hybrid models outshine their single-model counterparts in
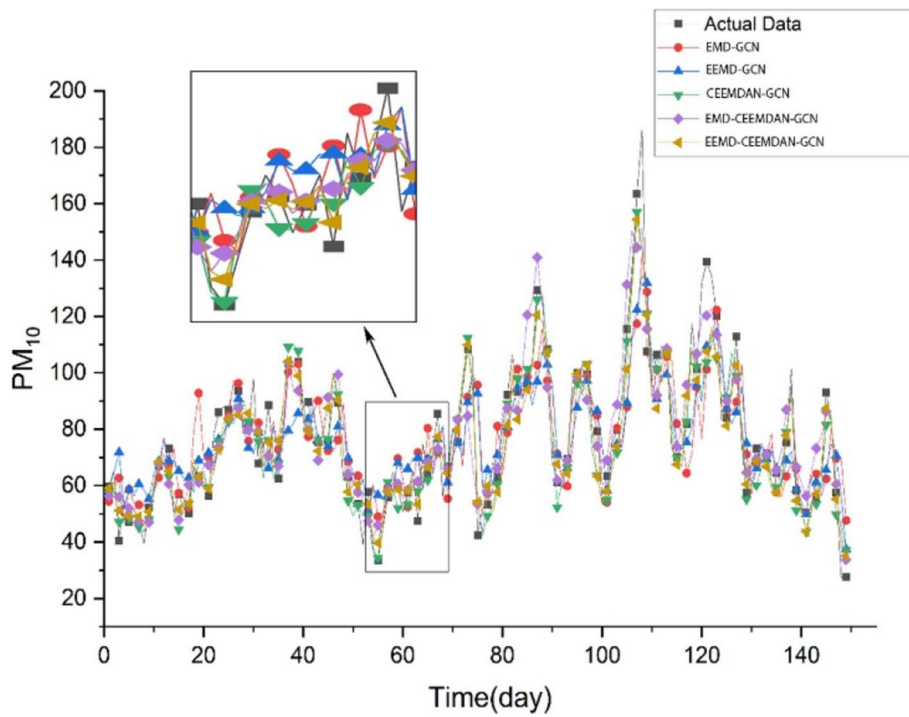
Bhatti *et al. Journal of Cloud Computing*   (2024) 13:65

Page 13 of 18

Fig. 5 Time series comparison of prediction results of 150 points observation of all the algorithms for the PM$_{10}$ dataset
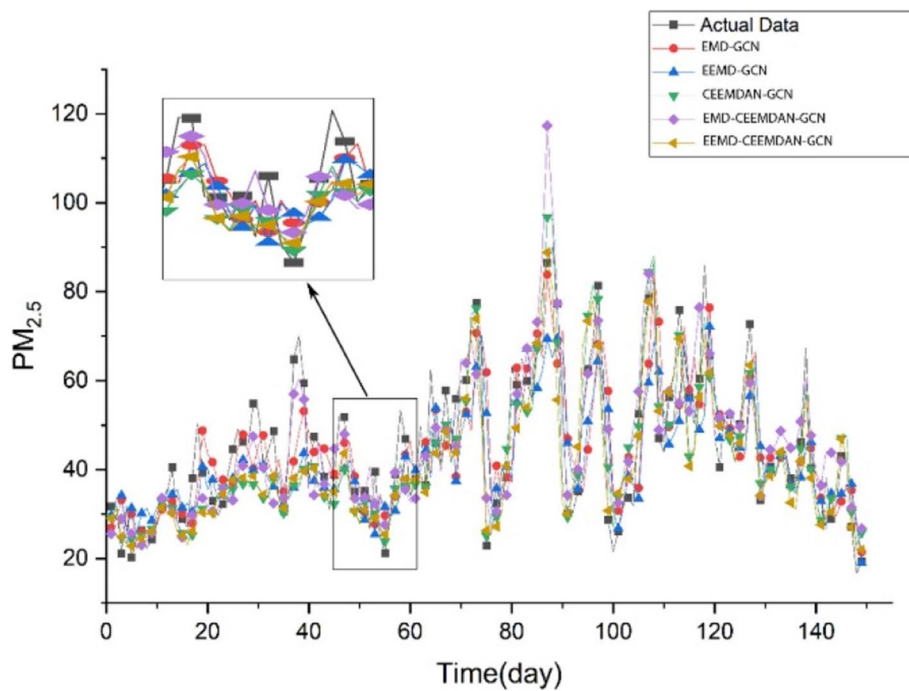

Fig. 6 Time series comparison of prediction results of 150 points observation of all the algorithms for the PM$_{2.5}$ dataset
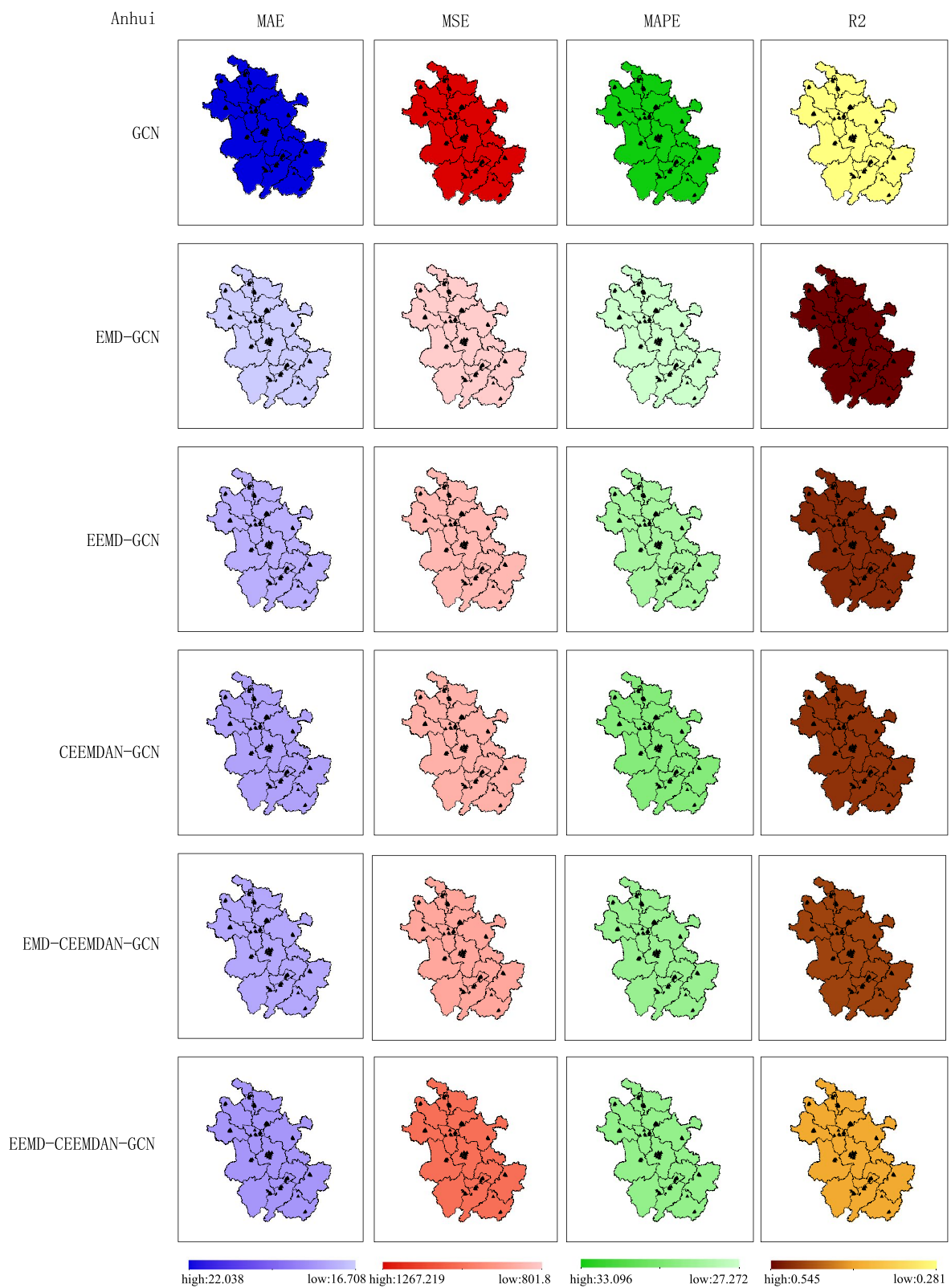
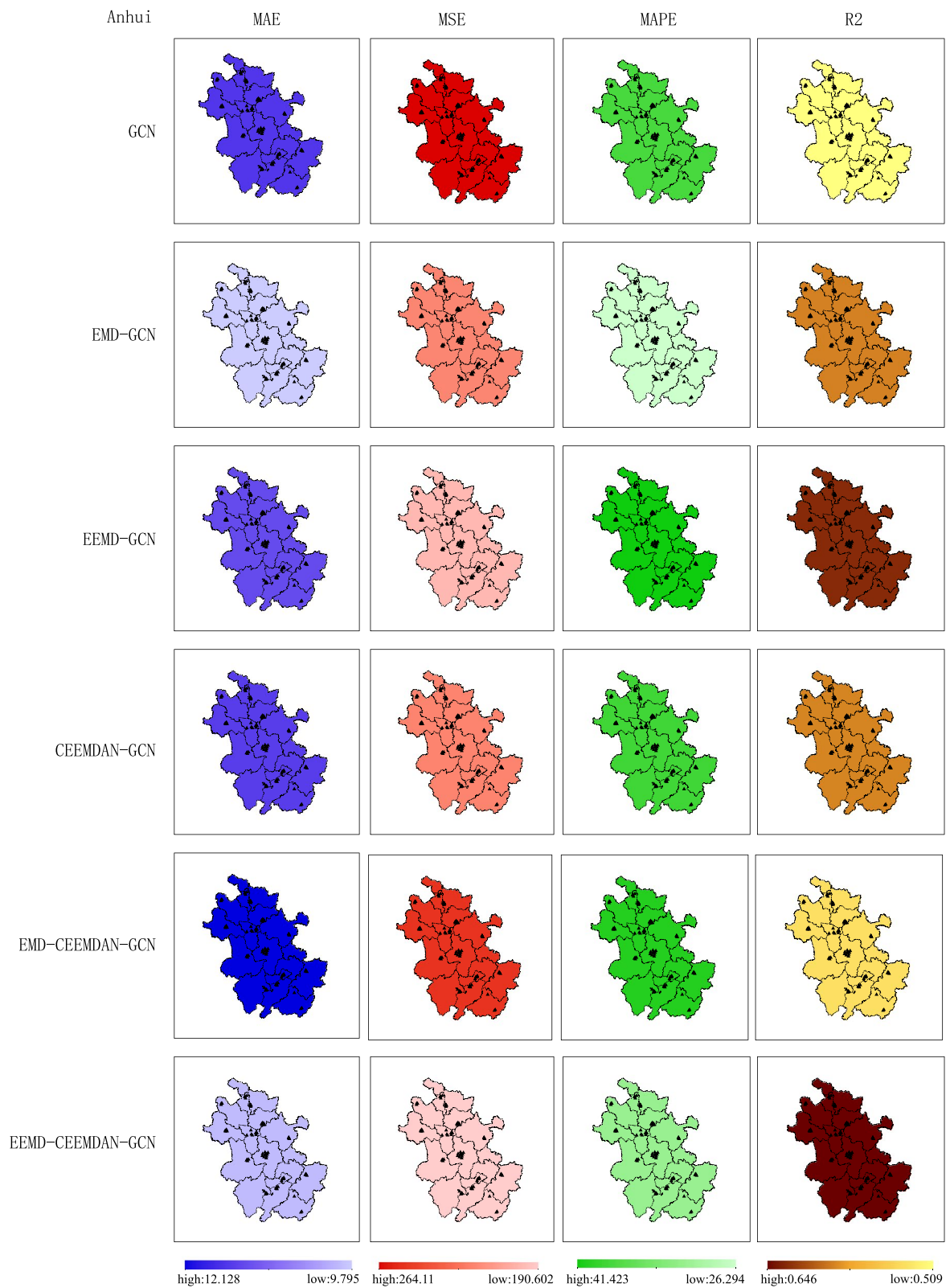**Fig. 7** Time series comparison of prediction results of 150 points observation of all the algorithms for the PM$_{10}$ dataset

Bhatti *et al. Journal of Cloud Computing*   (2024) 13:65

Page 15 of 18

Fig. 8 Time series comparison of prediction results of 150 points observation of all the algorithms for the PM$_{2.5}$ dataset

Bhatti *et al. Journal of Cloud Computing*   (2024) 13:65

Page 16 of 18

time series prediction. The amalgamation of multiple models seamlessly blends their respective strengths, enabling hybrid models to overcome the limitations of their constituents. In our quest for improved accuracy and computational efficiency, we delved into the realm of signal decomposition algorithms, employing them to disassemble distinct components of time series data. By selectively discarding extraneous feature components from these decomposed signals, we could merge the most valuable characteristics, balancing enhanced prediction accuracy and reduced computational complexity. Identifying and integrating these crucial components resulted in a more precise and simplified representation of the original time series data. This method bears particular significance when grappling with vast. Overall, EEMD-CEEMDAN-GCN offers improved decomposition accuracy, captures local and global features, effectively models-temporal dynamics, demonstrates robust predictive performance, and applies to multiple provinces. These benefits contribute to its effectiveness in analyzing and predicting air quality, supporting informed decision-making, and environmental management interventions.

Conclusion and future work

This research gives a thorough investigation of time series prediction utilizing the hybrid model EEMD-CEEMDAN-GCN, which blends deep learning with signal decomposition methods. Experimental findings using datasets on air quality in Chinese provinces show how well the suggested model works. The paper emphasizes how signal decomposition techniques like EEMD, CEEMDAN, and GCN may simplify time series data and boost prediction precision. The value of a multi-decomposition technique for comprehending and capturing underlying patterns in time series data is highlighted by the fact that combining different decomposition approaches improves prediction performance.

Addressing the unpredictability of air quality data and the scarcity of long-term historical data is a critical aspect of the research. The EEMD-CEEMDAN-GCN model is designed to mitigate these challenges through its unique combination of Ensemble Empirical Mode Decomposition (EEMD), Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN), and Graph Convolutional Network (GCN) components. The following sections elaborate on how the model effectively addresses these challenges, drawing comparisons with traditional methods.

Traditional methods often struggle with the non-linear and unpredictable nature of air quality data. Linear models and simple time-series approaches may overlook intricate patterns and fail to adapt to sudden changes in pollution levels, especially those reliant on statistical techniques, may struggle when faced with limited historical data. Their performance tends to degrade when there is a scarcity of examples to learn from, making it challenging to capture the complexity of air quality dynamics.

The EEMD-CEEMDAN-GCN hybrid model outperforms existing models in predictive performance. Its ability to fit and predict absolute values makes it practical for real-life scenarios, particularly in predicting air quality across different regions. This supports decision-makers in taking appropriate actions to mitigate pollution and improve public health. To account for the complex and dynamic nature of time series data, future research can explore incorporating additional factors into the prediction model, such as pollutant concentrations and weather factors. Integrating these factors can further enhance prediction accuracy, enabling more accurate forecasts of air pollution levels. Real-time performance is crucial in practical applications, and future work can focus on integrating online learning techniques with the proposed prediction model. Online learning allows for timely feedback and quick updates to model parameters, enabling more adaptive and responsive predictions, particularly in rapidly changing air quality conditions or extreme weather events.

The EEMD-CEEMDAN-GCN model, with its unique combination of signal decomposition, noise reduction, and graph-based learning, holds promise for a diverse range of applications. Concrete examples and case studies illustrate its potential impact in various fields:

> Implementing the EEMD-CEEMDAN-GCN model for real-time air quality prediction and management. Utilizing the model to analyze medical sensor data for early detection of environmental factors contributing to respiratory diseases. Applying the model to analyze climate data and understand the impact of air pollution on climate change variables. Integrating the model into urban planning to optimize infrastructure development considering air quality patterns.
>
> The EEMD-CEEMDAN-GCN model can be computationally intensive, particularly during the training phase, due to the iterative nature of ensemble decomposition and graph convolution operations. Future optimizations in parallel computing and model architecture may alleviate this limitation. While the model exhibits adaptability to limited historical data, substantial performance improvements could be achieved with larger and more diverse datasets. Future research should explore methods for transfer learning or domain adaptation to enhance the model's generalization capabilities with smaller

Bhatti *et al. Journal of Cloud Computing*   (2024) 13:65

Page 17 of 18

datasets. Integrate the model into comprehensive environmental monitoring systems, encompassing water quality, soil health, and biodiversity. Implementing the model to monitor and optimize industrial processes with potential environmental impacts. Applying the model to assess the financial impact of air quality on various industries.

The study suggests exploring other hybrid model combinations beyond signal decomposition and deep learning methods. Techniques like transfer learning, reinforcement learning, and online learning can be considered to design more advanced prediction models, pushing the boundaries of time series prediction accuracy. The proposed EEMD-CEEMDAN-GCN hybrid model extends beyond air quality prediction and can be applied to various time series data prediction scenarios in domains such as gold prices, wind speeds, and network traffic. The model's versatility opens up opportunities for application across different fields, demonstrating its broad potential and significance in addressing diverse prediction challenges.

Availability of data and materials
No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]School of Geography, Nanjing Normal University, Nanjing 210023, China. [2]Jiangsu Center for Collaborative Innovation in Geographical Information, Resource Development and Application, Nanjing 210023, China. [3]School of Information and Communication Engineering, Hainan University, Haikou 570100, China. [4]Guangdong-Hong Kong-Macao GBA New Generation Intelligent IoT Research Center, Shenzhen University, Shenzhen 518060, Guangdong, China.

References
1. Syuhada G, Akbar A, Hardiawan D, Pun V, Darmawan A, Heryati SHA, Siregar AYM, Kusuma RR, Driejana R, Ingole V et al (2023) Impacts of air pollution on health and cost of illness in Jakarta, Indonesia. Int J Environ Res Public Health 20:2916
2. Zhu L, El Khoudary S, Adibi J, Youk A, Talbott E (2022) Investigations of PM2.5 long-term exposure and subclinical atherosclerosis in women: an overview. 7:22–37. https://doi.org/10.17140/EPOJ-7-129
3. Guo Z, Miao Z, Guo F, Guo Y, Feng Y, Wu J, Zhang Y (2022) Parameter optimization of waste coal briquetting and particulate matter emissions test during combustion: a case study. Environ Pollut 294:118621
4. Dutta J, Roy S (2021) IndoorSense: context based indoor pollutant prediction using SARIMAX model. Multimed Tools Appl 80:19989–20018. https://doi.org/10.1007/s11042-021-10666-w
5. Lai K, Xu H, Sheng J, Huang Y (2023) Hour-by-hour prediction model of air pollutant concentration based on EIDW-informer—a case study of Taiyuan. Atmosphere 14:1274
6. Vignesh PP, Jiang JH, Kishore P (2023) Predicting PM$_{2.5}$ concentrations across USA using machine learning. Earth Space Sci 10:e2023EA002911
7. Mu G, Liao Z, Li J, Qin N, Yang Z (2023) IPSO-LSTM hybrid model for predicting online public opinion trends in emergencies. PLoS ONE 18(10):e0292677
8. Hakimi A, Monadjemi SA, Setayeshi S (2021) An introduction of a reward-based time-series forecasting model and its application in predicting the dynamic and complicated behavior of the earth rotation (Delta-T values). Appl Soft Comput 113:107920
9. Zhang R, Song H, Chen Q, Wang Y, Wang S et al (2022) Comparison of ARIMA and LSTM for prediction of hemorrhagic fever at different time scales in China. PLoS ONE 17(1):e0262009
10. Chen Y, Peng G, Zhu Z, Li S (2020) A novel deep learning method based on attention mechanism for bearing remaining useful life prediction. Appl Soft Comput 86:105919
11. Patra S (2017) Time series forecasting of air pollutant concentration levels using machine learning. Adv Comput Sci Inf Technol 4(5):280–284
12. Dua D, Graff C (2019) UCI machine learning repository. School of Information and Computer Science. University of California, Irvine
13. Bekkar A, Hssina B, Douzi S et al (2021) Air-pollution prediction in smart city, deep learning approach. J Big Data 8:161
14. Feng H, Zhang X (2023) A novel encoder-decoder model based on Autoformer for air quality index prediction. PLoS ONE 18(4):e0284293
15. Waseem KH, Mushtaq H, Abid F, Abu-Mahfouz AM, Shaikh A, Turan M, Rasheed J (2022) Forecasting of air quality using an optimized recurrent neural network. Processes 10:2117
16. Mirzavand Borujeni S, Arras L, Srinivasan V et al (2023) Explainable sequence-to-sequence GRU neural network for pollution forecasting. Sci Rep 13:9940
17. Qi Z, Wang T, Song G, Hu W, Li X, Zhang Z (2018) Deep air learning: interpolation, prediction, and feature analysis of fine-grained air quality. IEEE Trans Knowl Data Eng 30(12):2285–2297
18. Ahmed M, Shen Y, Ahmed M, Xiao Z, Cheng P, Ali N, Ghaffar A, Ali S (2022) AQE-Net: a deep learning model for estimating Air Quality of Karachi City from Mobile images. Remote Sens 14:5732
19. Masih A (2019) Machine learning algorithms in air quality modeling. Glob J Environ Sci Manag 5(4):515–534
20. Zhang L, Liu J, Feng Y et al (2023) PM2.5 concentration prediction using weighted CEEMDAN and improved LSTM neural network. Environ Sci Pollut Res 30:75104–75115
21. Chen S, Zheng L (2022) Complementary ensemble empirical mode decomposition and independent recurrent neural network model for predicting air quality index. Appl Soft Comput 131:109757
22. Zhao HR, Zhao YH, Guo S (2020) Short-term load forecasting based on complementary ensemble empirical mode decomposition and long short-term memory. Electr Power 53(06):1–8
23. H You, S Bai, R Wang, Z Li, S Xiang, F Huang, JS Mandeep (2022) New PSO-SVM Short-Term Wind Power Forecasting Algorithm Based on the CEEMDAN Model. JECE 2022. https://doi.org/10.1155/2022/7161445
24. Jaros R, Byrtus R, Dohnal J et al (2023) Advanced signal processing methods for condition monitoring. Arch Computat Methods Eng 30:1553–1577
25. Gao BX, Huang XQ, Shi JS, Tai YH, Zhang J (2020) Hourly forecasting of solar irradiance based on CEEMDAN and multi-strategy CNN-LSTM neural networks. Renew Energy 162:1665–1683

26. Luo J, Liang X, Guo Q, Zhang L, Bu X (2023) Combined improved CEEM-DAN and wavelet transform sea wave interference suppression. Remote Sens 15:2007
27. Chimmula VKR, Zhang L (2020) Time series forecasting of COVID-19 transmission in Canada using LSTM networks. Chaos Solit Fractals 135:109864
28. Sangiorgio M, Dercole F (2020) Robustness of LSTM neural networks for multi-step forecasting of chaotic time series. Chaos Solit Fractals 139:110045
29. Xiong Y, Zhao H (2019) Chaotic time series prediction based on long short-term memory neural networks. Sci Sin Phys Mechanic Astron 49(12):120501
30. Ghimire S, Deo RC, Casillas-Pérez D, Salcedo-Sanz S (2022) Improved complete ensemble empirical mode decomposition with adaptive noise deep residual model for short-term multi-step solar radiation prediction. Renew Energy 190:408–424
31. Li Y, Yu R, Shahabi C, Liu Y (2017) Diffusion convolutional recurrent neural network: data-driven traffic forecasting. arXiv:1707.01926
32. Li M, Zhu Z (2021) Spatial–temporal fusion graph neural networks for traffic flow forecasting. Proc AAAI Conf Artif Intell 35(5):4189–4196
33. Zhou F, Yang Q, Zhang K, Trajcevski G, Zhong T, Khokhar A (2020) Reinforced spatiotemporal attentive graph neural networks for traffic forecasting. IEEE Internet Things J 7(7):6414–6428
34. Wang X, Ma Y, Wang Y, Jin W, Wang X, Tang J, Jia C, Yu J (2020) Traffic flow prediction via spatial temporal graph neural network. Proc Web Conf 1082–1092
35. Zhang W, Liu H, Liu Y, Zhou J, Xiong H (2020) Semi-supervised hierarchical recurrent graph neural network for city-wide parking availability prediction. Proc AAAI Conf Artif Intell 34:1186–1193
36. Zhou H, Ren D, Xia H, Fan M, Yang X, Huang H (2021) AST-GNN: An attention-based spatio-temporal graph neural network for interactiona-ware pedestrian trajectory prediction. Neurocomputing 445:298–308
37. Mohamed A, Qian K, Elhoseiny M, Claudel C (2020) Social-STGCNN: A social spatio-temporal graph convolutional neural network for human trajectory prediction. Proc IEEE/CVF Conf Comput Vis Pattern Recognit (CVPR) 14424–14432
38. Han J, Liu H, Xiong H, Yang J (2023) Semi-supervised air quality forecasting via self-supervised hierarchical graph neural network. IEEE Trans Knowl Data Eng 35(5):5230–5243
39. Ram R, Venkatachalam Kv, Masud M, Abouhawwash M (2022) Air pollution prediction using dual graph convolution LSTM technique. Int Autom Soft Comput 33:1639–1652. https://doi.org/10.32604/iasc.2022.023962
40. Xu Z, Kang Y, Cao Y, Li Z (2021) Spatiotemporal graph convolution multi-fusion network for urban vehicle emission prediction. IEEE Trans Neural Netw Learn Syst 32(8):3342–3354
41. Khodayar M, Wang J (2019) Spatio-temporal graph deep neural network for short-term wind speed forecasting. IEEE Trans Sustain Energy 10(2):670–681
42. Huang Y, Yu J, Dai X, Huang Z, Li Y (2022) Air-quality prediction based on the EMD-IPSO-LSTM combination model. Sustain Times 14:1–18
43. Qu H, Zhang R (2022) Short-term mathematical prediction model of air quality based on CEEMD-ELM-PSO. 2022 IEEE International Conference on Electrical Engineering, Big Data and Algorithms (EEBDA), Changchun, China, pp. 227–232. https://doi.org/10.1109/EEBDA53927.2022.9744927
44. Wang DS, Wang HW, Lu KF, Peng ZR, Zhao J (2022) Regional prediction of ozone and fine particulate matter using diffusion convolutional recurrent neural network. Int J Environ Res Publ Health 19:3988
45. Li P, Zhang T, Jin Y (2023) A spatio-temporal graph convolutional network for air quality prediction. Sustainability 15:7624
46. Zhang JL, Che n F, Guo YN, Li XH (2020) Multi-graph convolutional network for short-term passenger flow forecasting in urban rail transit. IET Intell Transp Syst 14:1210–1217
47. Wu CL, He HD, Song RF, Zhu XH, Peng ZR, Fu QY, Pan J (2023) A hybrid deep learning model for regional O3 and NO2 concentrations prediction based on spatiotemporal dependencies in air quality monitoring network. Environ Pollut 320:121075

Publisher's Note

Mughair Aslam Bhatti   (Fellow, IEEE) and all authors may include biographies if the publication allows. Biographies are often not included in conference-related papers. Please check the Information for Authors to confirm. Author photos should be current, professional images of the head and shoulders. The first paragraph may contain a place and/or date of birth (list place, then date). Next, the author's educational background is listed. The degrees should be listed with the type of degree in what field, which institution, city, state, and country, and year the degree was earned. The author's major field of study should be lowercase.The second paragraph uses the preferred third person pronoun (he, she, they, etc.) and not the author's last name. It lists military and work experience, including summer and fellowship jobs. Job titles are capitalized. The current job must have a location; previous positions may be listed without one. Information concerning previous publications may be included. The format for listing publishers of a book within the biography is: *Title of Book* (publisher name, year) similar to a reference. Current and previous research interests end the paragraph.

The third paragraph begins with the author's preferred title and last name (e.g., Dr. Smith, Prof. Jones, Mr. Kajor, Ms. Hunter, Mx. Riley). List any memberships in professional societies other than the IEEE. Finally, list any awards and work for IEEE committees and publications.

Zhiyao Song   Photograph and biography not available at the time of publication.

Zhiyao Song   Photograph and biography not available at the time of publication.