IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 13 July 2024, accepted 30 July 2024, date of publication 9 August 2024, date of current version 12 September 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3441109

RESEARCH ARTICLE

Future Air Quality Prediction Using Long Short-Term Memory Based on Hyper Heuristic Multi-Chain Model

KALYAN CHATTERJEE[1], (Member, IEEE), SAMLA SURAJ KUMAR[2], (Student Member, IEEE), RAMAGIRI PRAVEEN KUMAR[1], ANJAN BANDYOPADHYAY[3], SUJATA SWAIN[3], SAURAV MALLIK[4], (Member, IEEE), AMAL AL-RASHEED[5], MOHAMED ABBAS[6], AND BEN OTHMAN SOUFIENE[7]

[1]Department of Computer Science and Engineering, Nalla Malla Reddy Engineering College, Hyderabad 500088, India
[2]Department of Artificial Intelligence and Machine Learning, Nalla Malla Reddy Engineering College, Hyderabad 500088, India
[3]School of Computer Science and Engineering, Kalinga Institute of Industrial Technology, Bhubaneswar 751024, India
[4]Department of Environmental Health, Harvard T. H. Chan School of Public Health, Boston, MA 02115, USA
[5]Department of Information Systems, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia
[6]Electrical Engineering Department, College of Engineering, King Khalid University, Abha 61421, Saudi Arabia
[7]PRINCE Laboratory Research, ISITcom, Hammam Sousse, University of Sousse, Sousse 4000, Tunisia

Corresponding author: Ben Othman Soufiene (soufiene.benothman@isim.rnu.tn)

This work was supported by Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2024R235), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. The authors extend their appreciation to the Deanship of Research and Graduate Studies at King Khalid University for funding this work through Large Research Project under grant number RGP2/549/45.

ABSTRACT Air pollution is a critical global concern, demanding precise air quality forecasting to mitigate its severe consequences. Our study introduces Future Air Quality Prediction using Long Short-Term Memory based on Hyper Heuristic Multi-Chain Model (H^2MCM) to project future air quality, considering various meteorological factors (MFs) and pollution-related variables like atmospheric pressure, temperature, humidity, and wind patterns. Leveraging 12 units of Long Short-Term Memory neural networks (LSTMs), H^2MCM accurately predicts forthcoming air pollutants (APs) concentrations such as particulate matter with diameter 2.5 μm ($PM_{2.5}$), carbon monoxide (CO), and nitrogen dioxide (NO_2). Additionally, it accounts for spatiotemporal correlations between these APs and MFs, which significantly influence the air quality prediction for the next immediate time interval. H^2MCM utilizes a multi-chain mechanism, employing *1-hour prediction model*s to forecast air quality hourly, enabling approximations for the next 12 hours. Also, for an efficient model selection, Akaike Information Criterion (AIC), Schwarz Bayesian Information Criterion (SBIC), Hannan-Quinn Information Criterion (HQIC), and corrected AIC (AICc) tools are used based on their ability to balance model fit and complexity. Furthermore, it demonstrates the ability to enhance the performance of any predictor. Experimental results substantiate H^2MCM's superiority over various models, including the Support Vector Regressor (SVR), Multi-Layer Perceptron (MLP), Recurrent Air Quality Predictor (RAQP), and Valchogianni models. H^2MCM achieves impressive up to 75% better accuracy and consistency compared to SVR, 60% better than MLP, 38% better than RAQP, and 70% better than Valchogianni models.

INDEX TERMS Air quality, air pollutant concentrations (APCs), deep learning (DL), heuristic, machine learning (ML), meteorological factors (MFs), multi-chain, regressors.

I. INTRODUCTION

This urban development and industrialization have led to a severe air pollution problem in urban areas, posing

The associate editor coordinating the review of this manuscript and approving it for publication was Diego Oliva.

significant risks to human health and the environment and impacting global economic systems. According to the World Health Organization (WHO), excessive air pollutants cause approximately 4.2 million deaths annually, with 9 out of 10 people suffering from breathing problems due to these pollutants [1]. Therefore, there is an urgent need to develop an efficient technique for predicting air quality in the coming hours to aid in environmental cleanup.

Assessing the impact of APs on air quality estimation presents a significant challenge due to their non-linear and dynamic nature in real-world processes [10]. The conventional chemical transport model (CTM) [15] for air quality estimation requires vast data. However, given the demand for advanced models capable of handling non-linearity in real-world processes, deep learning-based regression algorithms like LSTM, MLP and Artificial Neural Networks (ANN) are increasingly employed to map non-linear inputs (i.e., air pollutant concentrations and meteorological factors) to predicted outputs (i.e., future air quality estimations). Several studies [5], [13] have presented deep learning-based non-linear air pollution models, while others [8], [9] have successfully applied machine learning-based regressors to predict future air quality. Section VI of this study presents a performance analysis of these models.

This study investigates various regression techniques using ML/DL to predict future air quality. Experimental results indicate that the ML-based support vector regression (SVR) [24] technique performs well for short time intervals due to strong correlations among air pollution components (APCs) and meteorological factors (MFs) [10]. However, this correlation weakens as the time intervals increase, rendering SVR [24] unreliable for long-term predictions. Consequently, the focus shifts to the DL-based LSTM regression technique.

The study's notable contributions are as follows:

1) Introducing the *heuristic per hour mechanism* and *noise injection* to build the required time-series big dataset for future air quality prediction.
2) Developing 12 *1-hr Prediction Models* to create the multi-chain H^2MCM model. Each *1-hr Prediction Model* is needed to create the *chain* architecture of the H^2MCM model.
3) Defining short-, medium-, and long-term predictors by assessing the internal spatiotemporal correlation between APCs and MFs.
4) Combining the heuristic multi-chain H^2MCM model with the LSTM regressor to capture non-linear relationships among APCs and MFs.

To construct the big dataset, random sampling techniques are used, and noise/outlier sets [14] are added to the original dataset [4], [19] as described in subsection VI-A. Manual injection of noise/outlier sets serves two purposes: enhancing the capacity of the utilized regressors and improving prediction accuracy and consistency over extended time intervals.

The remaining structure of this study is organized as follows: Firstly, the relevant literature survey is presented in Section II. Following that, in Section III, the system model & problem formulation are introduced. The next section, Section IV, covers predictors based on ML/DL. The proposed H^2MCM model is detailed in Section V. Subsequently, in Section VI, the performance evaluation is conducted, comparing the H^2MCM model with state-of-the-art methodologies. Section VII described the discussion of this study. Finally, the study concludes and discusses future work in Section VIII.

II. LITERATURE SURVEY

The development of algorithms for air quality prediction has evolved significantly, driven by the need to mitigate the harmful effects of air pollution through accurate forecasting. Traditional theory-based approaches, which rely on complex mathematical expressions, often fall short in addressing the intricacies of APCs, the variability in emission rates, and the lack of comprehensive data on pollution sources. In response, ML/DL algorithms have emerged as powerful tools, offering enhanced capabilities in modeling and predicting air quality dynamics.

A. AIR QUALITY PREDICTION USING ML/DL ALGORITHMS

Intelligent prediction systems play a vital role in minimizing the harmful effects of air pollution through short-, medium-, and long-term forecasting. However, the intricate nature of APCs, insufficient data on pollution sources, and constantly changing emission rates pose significant obstacles to creating these systems. ML/DL algorithms have demonstrated tremendous success in tackling the intricate nature of APCs, along with insufficient data on pollution sources and constantly changing emission rates, which pose significant obstacles in creating these systems compared to theory-based approaches relying on complex mathematical expressions.

B. EARLY DEVELOPMENTS OF NEURAL NETWORKS AND TIME-SERIES ANALYSIS

The introduction of neural networks (NNs) and ML/DL methods marked a significant breakthrough in the field of air quality prediction. These methods, particularly artificial neural networks (ANNs), were employed to evaluate levels of various air pollutants and categorize air quality. Initial studies demonstrated that NNs could effectively handle the nonlinear and complex relationships inherent in air pollution data [34], [35], [33].

Recurrent Neural Networks (RNNs) [39], with their ability to link and train multiple layers of neurons with predefined weights, became instrumental in analyzing time-series data. RNNs excel in capturing temporal dependencies, making them suitable for forecasting future air quality based on historical pollutant levels. Research leveraging RNNs has shown significant promise in improving the accuracy of air quality predictions by considering past data trends and patterns [36], [37], [38].

TABLE 1. Existing research works on air pollution prediction using various ML- & DL-based regression algorithms.

Authors	Problem addressed	Regressor used	Dataset	Proposed Solution
Yang et. al. [5]	Constructed a future air quality prediction model using MFs and various APCs.	LSTM & GRU Models	Used public dataset from the Harvard Dataverse [16]	i) Exploited the Shapley Additive Explanation (SHAP) method for analyzing future air quality. ii) Exploited the correlation among various participating air pollutants with meteorological factors.
Gu et. al. [2]	Developed a precise model for predicting air pollution levels during large gatherings and demonstrations.	Multitask temporal support vector regressor (TSVR)	Used their own dataset	i) Exploited air pollution in mass rallies. ii) Established a temporal weighting matrix for use in air pollution forecasting.
Liu et. al. [33]	Predict the future urban air quality.	RNN with GeoSOM algorithm	Real air quality data	i) Exploits patiotemporal RNN with GeoSOM to predict future urban $PM_{2.5}$ concentration during next 24 hours.
Chiang & Horng [31]	Built a hybrid framework for daily-basis $PM_{2.5}$ prediction in time-series data.	GRU	Collected from Taiwan Environmental Protection Administration (TWEPA).	i) Utilized location-centric spatial features along with daily-based temporal features to achieve precise future predictions.
Saiohai et al. [3]	Developed an early warning system for predicting air quality.	Weka algorithm & vertical MFs	Own dataset.	i) Forecast local $PM_{2.5}$ levels and episodes of excessive $PM_{2.5}$ concentrations utilizing vertical MFs.

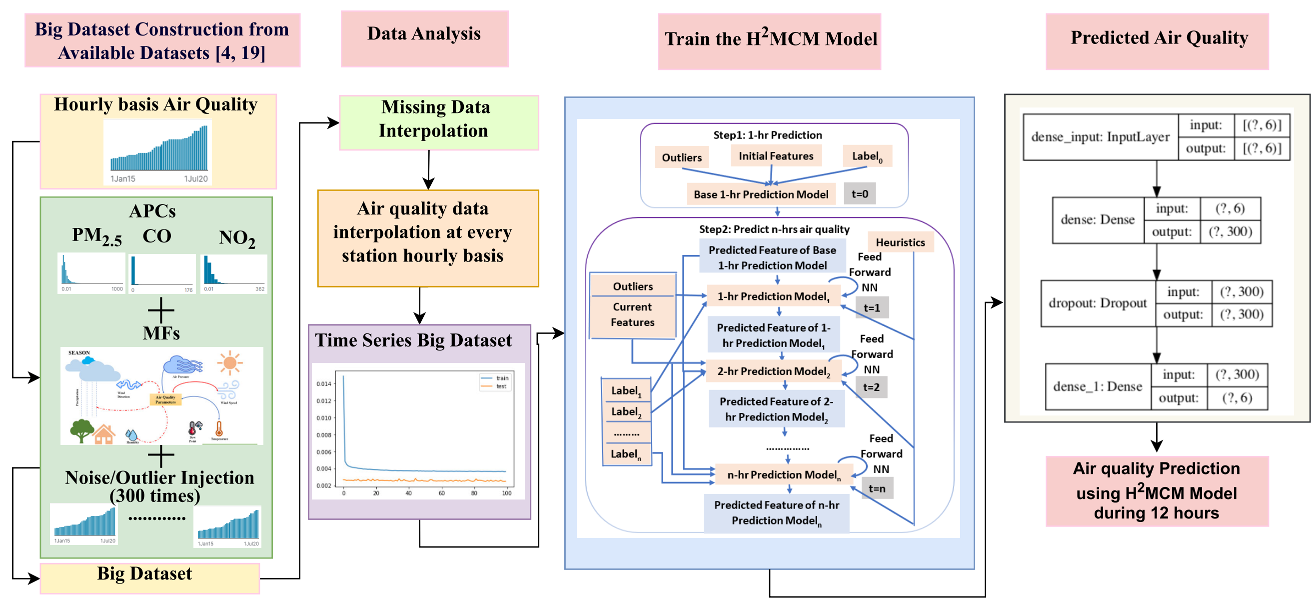


FIGURE 1. Architecture of the proposed H^2MCM model.

C. ADVANCEMENTS IN URBAN AIR QUALITY PREDICTION MODELS

Urban air quality prediction models can be broadly categorized into theory-driven physical processes and data-driven statistical approaches. Each approach has its computational demands, but data-driven methods, particularly those using ML/DL techniques, have proven to be more adaptable and cost-effective as emphasized in Table 1.

The rapid advancement in data collection technologies has enabled the accumulation of vast amounts of urban big data, which in turn has fueled the development of sophisticated data-driven models [6], [23]. The Long Short-Term Memory (LSTM) networks [7] and Gated Recurrent Units (GRUs) [31] have shown exceptional capabilities in managing sequential time-series data, a common requirement in air quality prediction. These models excel at capturing long-term dependencies and have been widely adopted in air pollution research [7], [11], [31], [32].

Researchers have also integrated LSTM networks with Convolutional Neural Networks (CNNs) and Graph Convolutional Networks (GCNs) to capture both temporal and spatial patterns in air quality data [11], [32]. This hybrid approach has enhanced the accuracy and robustness of predictions by leveraging spatial correlations among monitoring stations [29], [30]. Also, researchers have effectively combined LSTM with convolutional neural networks (CNN) [30] and graph CNN [29] to incorporate spatial patterns identified in monitoring station data. Additionally, the use of neural attention networks has further improved model performance by identifying and focusing on the most relevant features in the data [27], [28].

D. RECENT TRENDS IN FUSION AND ENSEMBLE STRATEGIES

In recent years, there has been a shift towards using fusion and ensemble strategies in DL models to improve long-term air quality predictions. These strategies combine multiple models to capture the spatiotemporal characteristics of air pollution and weather datasets more effectively. By integrating diverse models, researchers can enhance prediction accuracy and reliability, addressing the challenges posed by incomplete or noisy data [25], [26].

Despite these advancements, predicting long-term air pollution remains a formidable challenge. The limited availability of comprehensive and high-quality air pollution data, along with the inherent noise and missing values, continues to hinder the development of highly accurate models. Ongoing research focuses on refining these models, improving data preprocessing techniques, and exploring novel architectures to overcome these limitations.

Therefore, the development of algorithms for air quality prediction has progressed from basic neural networks to sophisticated ML/DL models capable of handling complex temporal and spatial dependencies. The integration of advanced techniques such as LSTM, GRU, CNN, GCN, and attention mechanisms has significantly improved the accuracy and robustness of air quality forecasts. However, challenges remain, particularly in long-term prediction, due to data limitations. Continued research and innovation in this field are essential to further enhance the predictive capabilities and reliability of air quality prediction models.

III. SYSTEM MODEL & PROBLEM FORMULATION

The expansion of power systems, networks, electric vehicles, and advanced high-speed technologies has been identified as a major factor contributing to the continuous increase in air, water, and food pollution. Among these types of pollution, air pollution has gained considerable attention from governments and the public due to its significant global impacts. As a result, there is an urgent need for an accurate model capable of monitoring and predicting air quality to protect the planet. To meet this demand, ML [18], and DL [20], [21], [22] have emerged as powerful tools, as evident from their increasingly widespread use.

A. SYSTEM MODEL

System architecture of the H^2MCM model is illustrated in Figure 1. This model comprises a multi-chain architecture with 12 units of *1-hr Prediction Model* designed to forecast the APCs 12 hours ahead from the present time, as detailed in Table 2. The creation of each *1-hr Prediction Model* involves employing a per hour heuristic multi-chain mechanism.

Table 2 shows that the accuracy and consistency of the H^2MCM model display improvement with an increasing number of *1-hr Prediction Model*s. Each *1-hr Prediction Model* exhibits higher values for *PLCC* and *RMSE*, indicating enhanced performance, while lower values of the *Delta Rule* signify reduced network errors. However, beyond 12 units of *1-hr Prediction Model*s, the accuracy improvement becomes negligible. Adding 13 units results in only a 0.5% accuracy improvement; subsequent units do not yield significant enhancements. Moreover, with the inclusion of more units, the computational complexity of the model continues to rise. Consequently, we conclude that the optimal choice is to employ 12 units of *1-hr Prediction Model* for constructing our multi-chain H^2MCM model.

TABLE 2. Number of ***1-hr Prediction Model*** for the construction of H^2MCM model.

1-hr Prediction Model Units	Accuracy (PLCC)	Consistency (RMSE)	Error (Delta-Rule)
1	40%	30%	80%
2	55%	45%	70%
3	68%	58%	55%
4	80%	75%	48%
5	92%	88%	35%
6	93.5%	90%	30%
7	94%	91%	32%
8	95%	92%	30%
9	95.95%	92.36%	30.25%
10	96.5%	93%	30%
11	96.68%	93%	29.69%
12	**98.59%**	**94.28%**	**27.23%**
13	99%	90%	30%
14	99.23%	90%	30%
15	99.23%	90%	30%
16	99.23%	90%	30%

B. MODEL SELECTION

Model selection is crucial in ensuring that the chosen model effectively represents the data and provides accurate future predictions. In this context, information criteria such as Akaike Information Criterion (AIC), Schwarz Bayesian Information Criterion (SBIC), Hannan-Quinn Information Criterion (HQIC), and corrected AIC (AICc) are valuable tools for comparing and selecting the most appropriate model. These information criteria are used to strike a balance between model fit and model complexity. They penalize overly complex models, encouraging the selection of models that are both parsimonious and predictive. They can significantly aid in the decision-making process when choosing among competing models. Table 3 presents selection criteria of our proposed DL-based future air quality prediction model.

Lower values for AIC, SBIC, HQIC, and AICc indicate better model fit and more fabulous parsimony. In the context of these criteria, a lower value represents a more favorable trade-off between model fit and complexity.

LSTM-based H^2MCM has the lowest AIC, SBIC, HQIC, and AICc values among all models, suggesting that it is the most parsimonious and predictive model according to these criteria.

RAQP-based H^2MCM also performs well, with the second-lowest values across all criteria. It is a strong contender and may be considered for further analysis.

SVR-based H^2MCM, MLP-based H^2MCM, and Vlachogianni-based H^2MCM have higher AIC, SBIC, HQIC,

TABLE 3. Model comparison using various information criteria.

Model	AIC	SBIC	HQIC	AICc
SVR-based H^2MCM	1200.5	1250.3	1225.6	1205.7
RAQP-based H^2MCM	1198.2	1247.9	1220.4	1199.8
MLP-based H^2MCM	1202.6	1253.4	1228.9	1205.5
Vlachogianni-based H^2MCM	1199.4	1249.0	1219.5	1198.3
LSTM-based H^2MCM (Proposed)	**1170.4**	**1200.0**	**1209.5**	**1180.3**

and AICc values, indicating a relatively poorer trade-off between model fit and complexity. These models may be considered less parsimonious according to these criteria.

It is important to note that the specific interpretation of these criteria can vary depending on the context and the complexity of the compared models. The model with the lowest values across all four criteria is often considered the best trade-off between model complexity and fit. Therefore, from Table 3, we observed the lowest values across all four criteria for LSTM-based H^2MCM and hence, select LSTM-based H^2MCM to design our proposed future prediction model:H^2MCM.

C. PROBLEM FORMULATION

This study aims to develop a dependable and accurate air pollution model capable of predicting the hourly levels of APCs and MFs for the next 12 hours. To achieve this goal, we have defined our objective function using the following Equation (1).

$$(f_i : 1 \leftarrow n) \xrightarrow{Hyper\ heuristic\ multi-chain\ LSTM} (g_i : 1 \leftarrow n) \tag{1}$$

Here, f and g are the real functions which represent the input and output. n is the total number of samples. i is the total number of iterations.

IV. ML- AND DL-BASED PREDICTION

Within this section, we have presented two regression algorithms to tackle the issue of predicting future air quality. Specifically, SVR [24] represents the ML-based prediction approach, while LSTM [7] represents a DL-based prediction approach.

A. SUPPORT VECTOR REGRESSOR (SVR)

The support vector regression (SVR) approach is an extension of the support vector machine (SVM) [24]. SVR [24] is well-suited for handling non-linear data and, in this context, is used to formulate the objective function for the ML-based prediction approach.

Let, $DS_{Tr} = \{(x_i, y_i)\}$ be the training samples, where, $i \in \mathbb{R}$, $x_i = [x_1^1, x_2^2, x_3^3, \ldots\ldots, x_n^n] \in \mathbb{R}^n$ represents the i^{th} feature input and $y_i \in \mathbb{R}$ denotes the i^{th} real target output.

The forthcoming Equation (2) illustrates the general form of SVR using a hyperplane function.

$$H(x_i) =< \upsilon, X(x_i) > +\psi \tag{2}$$

Here, the $<, , >$ operator denotes the inner product, while $X(, .)$, defines a non-linear function. The variables υ denotes feature inputs and ψ represent the target output parameters, respectively. The value of H can be normalized by $\min(, |\upsilon|^2),$. To incorporate margin of errors for SVR, two variables, namely a_1 and a_2, are used. Based on the above discussion, we can represent our objective function as a concave optimization problem, as shown in Equation (3):

$$min \left(|\upsilon|^2 + c * \sum_{j=1}^{n} (a_1 + a_2) \right) \xrightarrow{e} \begin{cases} H - y_i \leq e + a_1 \\ y_i - H \leq e + a_2 \\ a_1, a_2 \geq 0 \in a_j \end{cases} \tag{3}$$

Here, the variable e signifies computational errors, while c represents the regularization constant, whose value must be greater than 0. In the absence of regularization, the objective function is minimized by the variables a_1 and a_2, where $j = 1, 2, \ldots., m$. To mitigate the issues of under- and over-fitting, we set $c = |\upsilon|^2/2$, and aim to minimize $e = c*\sum_{j=1}^{n}(a_1+a_2)$.

To achieve an optimal learned solution, we also take into account the kernel function. The kernel function is expressed by the following Equation (4):

$$K(x_i, x_j) =< g(x_i), g(x_j) > \tag{4}$$

Here, the function g maps X ($x_i \in X \in R$) to the hyperplane and generates Y ($y_i \in Y \in R$). The sigmoid kernel ($K(x_i, x_j) = tanh(\gamma * x_i^T * x_j + r)$) is used to establish ML regression models. The hyperparameters γ and r are determined through training samples.

To predict air quality using an ML-based regressor, we use the SVR method. This method uses the initial values of MFs and APCs (*Base 1-hr Prediction Model* at time t=0) to forecast air quality from t=1 hour to t=12 hours.

Based on the findings in Table 4, it is evident that the SVR regressor exhibits higher average PLCC values and lower average RMSE values for shorter time intervals. However, as the prediction horizon extends beyond 5 hours, the SVR's performance deteriorates, indicated by a sharp drop in average PLCC and an increase in average RMSE. This decline is attributed to weak correlations among the involved air contaminants and MFs over longer time periods. Consequently, we deduce that the ML-based SVR approach is more suitable as a short-term predictor, prompting us to explore the DL-based LSTM technique for improved prediction accuracy.

B. LONG SHORT-TERM MEMORY (LSTM)

The Long Short-Term Memory (LSTM) network [7] is a type of recurrent neural network (RNN) [39] that is well-suited for time-series prediction problems. Unlike traditional RNNs, LSTMs can effectively capture long-term dependencies in

TABLE 4. Compared ML-based SVR Model DL-based LSTM Model.

Models	t=1hr	t=2hr	t=3hr	t=4hr	t=5hr	t=6hr	t=7hr	t=8hr	t=9hr	t=10hr	t=11hr	t=12hr
SVR (Avg. PLCC)	0.899	0.906	0.921	0.925	0.931	0.721	0.707	0.699	0.697	0.693	0.687	0.685
LSTM (Avg. PLCC)	0.900	0.909	0.917	0.923	0.923	0.923	0.935	0.731	0.732	0.726	0.719	0.715
SVR (Avg. RMSE)	0.035	0.039	0.047	0.051	0.055	0.062	0.064	0.072	0.075	0.076	0.077	0.077
LSTM (Avg. RMSE)	0.025	0.025	0.032	0.032	0.036	0.039	0.042	0.041	0.041	0.052	0.053	0.053

sequential data due to their unique architecture, which includes memory cells that can maintain information over extended periods.

An LSTM network is composed of a sequence of LSTM cells. Each cell in the LSTM network processes an input sequence, x_i, to produce an output, h_i, and maintains a cell state, c_i. The LSTM cell consists of three gates: the input gate, the forget gate, and the output gate, which regulate the flow of information through the cell.

The equations governing the operations of an LSTM cell are as follows:

1) FORGET GATE

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{5}$$

Here, f_t is the forget gate activation, σ is the sigmoid function, W_f is the weight matrix, h_{t-1} is the previous hidden state, x_t is the current input, and b_f is the bias.

2) INPUT GATE

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{6}$$

$$\tilde{c}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \tag{7}$$

Here, i_t is the input gate activation, $\tilde{c}_t$ is the candidate cell state, W_i and W_c are weight matrices, and b_i and b_c are biases.

3) CELL STATE UPDATE

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \tag{8}$$

Here, c_t is the updated cell state, and $\odot$ denotes element-wise multiplication.

4) OUTPUT GATE

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{9}$$

$$h_t = o_t \odot \tanh(c_t) \tag{10}$$

Here, o_t is the output gate activation, h_t is the hidden state, W_o is the weight matrix, and b_o is the bias.

The LSTM network leverages these gates to control the flow of information, thereby addressing the vanishing gradient problem commonly encountered in traditional RNNs.

To predict air quality using an LSTM-based approach, the initial values of meteorological factors (MFs) and air pollution components (APCs) are used as inputs to the LSTM network (*Base 1-hr Prediction Model* at time t=0). The LSTM network is trained to forecast air quality from t=1 hour to t=12 hours.

By incorporating LSTM networks, which are adept at capturing long-term dependencies and temporal patterns, we aim to achieve improved prediction accuracy over longer time horizons. The performance of the LSTM-based model is evaluated using metrics such as the PLCC and RMSE. Based on our findings, the LSTM approach exhibits higher PLCC values and lower RMSE values compared to traditional ML models, particularly for longer prediction intervals, demonstrating its suitability for medium- and long-term air quality forecasting.

V. PROPOSED LSTM-BASED HYPER HEURISTIC MULTI-CHAIN MODEL (H^2MCM)

The H^2MCM model is our proposed solution for medium- and long-term air quality prediction. To achieve the required prediction accuracy, we develop *1-hr Prediction Models* and use an error computation mechanism based on transformer-based multi-chain hyper-heuristic rules. This approach is taken because integrating transformer models directly into H^2MCM has certain limitations. Therefore, to leverage the accuracy advantages of transformer models in time-series air quality data, this study formulated the transformer-based multi-chain hyper-heuristic rules.

A. LIMITATIONS OF INTEGRATING TRANSFORMER MODELS WITH H^2MCM

1) **Computational Complexity:**
   a) **Transformers:** Transformer models, such as those discussed in [40] and [41], are computationally intensive due to their full attention mechanism. This mechanism exhibits quadratic complexity relative to the sequence length, leading to high memory consumption and slower training times, particularly for long sequences.
   b) **H^2MCM (LSTM-based): LSTM models are generally more efficient at handling long sequences because they process sequences sequentially and maintain a manageable computational complexity.**
2) **Hyperparameter Tuning:**
   a) **Transformers:** Transformers require careful tuning of many hyperparameters, such as the number of layers, attention heads, and size of the feed-forward network. This makes the training process more complex and time-consuming.
   b) **H^2MCM (LSTM-based):** While LSTM models also require hyperparameter tuning, the process is relatively more straightforward than

transformers. The primary hyperparameters include the number of LSTM units, the number of layers, and the dropout rate.

3) **Interpretability:**

a) **Transformers:** Due to their complex architecture, transformer models, such as those discussed in [40]and [41], can be less interpretable than LSTM models. Understanding and explaining transformers' internal workings and decision-making processes can be challenging.

b) **H^2MCM (LSTM-based):** LSTM models, while still complex, offer more straightforward interpretability through their gate mechanisms (input, forget, and output gates) and the sequential processing of data.

4) **Spatiotemporal Correlations:**

a) **Transformers:** While transformers handle spatial dependencies well due to their attention mechanism, capturing temporal correlations requires additional architectural modifications, such as incorporating temporal embeddings.

b) **H^2MCM (LSTM-based):** LSTM models are inherently designed to capture temporal dependencies, making them well-suited for time series data like APCs and MFs.

Thus, while transformer models, such as those discussed in [40] and [41], offer advanced capabilities and have shown superior performance in various dynamic systems, their integration into the H^2MCM model presents several challenges. These include higher computational complexity, more significant data requirements, and the need for extensive hyperparameter tuning. On the other hand, the LSTM-based H^2MCM is well-suited for handling temporal dependencies in time series data, offering a more efficient and interpretable approach with lower computational demands. However, recognizing the strengths of transformer models, such as their ability to capture global dependencies and handle spatial correlations effectively, we have designed our multi-chain hyper-heuristic rules based on transformer models.

B. TRANSFORMER-BASED MULTI-CHAIN HYPER-HEURISTIC RULES

Let M be the total number of chains. Each chain i has a set of transformer hyperparameters denoted by H_i, where $i = 1, 2, \ldots, M$.

The transformer-based objective function for the hyper-heuristic optimization is defined as:

$$\text{Objective Function: } \min(f(H)) = H \in H_1, H_2, \ldots, H_M \tag{11}$$

Here, H is the set of all transformer hyperparameters across all chains.

The solution space for the multi-chain hyper-heuristic is the union of all chain-specific solution spaces:

$$\text{Solution Space:} \quad S = \bigcup_{i=1}^{M} H_i \tag{12}$$

The optimization process aims to find the best combination of transformer hyperparameters that yield the optimal performance for the proposed H^2MCM model.

The transformer-based multi-chain hyper-heuristic leverages the power of transformer networks to capture spatiotemporal dependencies in the air quality time-series data, and the hyper-heuristic optimization explores the solution space to find the optimal transformer hyperparameters for each chain.

Therefore, by maintaining these multi-chain heuristic rules, our proposed H^2MCM model can enhance the accuracy and consistency of air quality prediction. Algorithm 1 outlines the proposed transformer-based multi-chain hyper-heuristic rules to develop the H^2MCM model.

Algorithm 1 Proposed Transformer-Based Multi-Chain Hyper-Heuristic Rules

Initialize the total number of chains M
Initialize the maximum number of iterations N_{iter}
Initialize the set of transformer hyperparameters for each chain: $H_1, H_2, \ldots, H_M$
Initialize the best hyperparameters: H
Initialize the best objective function value: $f = \infty$
for $i \leftarrow 1$ to M **do**
  for $j \leftarrow 1$ to N_{iter} **do**
    Randomly generate transformer hyperparameters H_i
    Train transformer model with hyperparameters H_i on training data
    Evaluate the objective function value: $f(H_i)$
    if $f(H_i) < f$ **then**
      Update the best hyperparameters: $H \leftarrow H_i$
      Update the best objective function value: $f^* \leftarrow f(H_i)$
    end if
  end for
end for
Output: The best set of transformer hyperparameters H^* that yield the optimal performance for air quality prediction.

C. TRAINING OF THE H^2MCM MODEL

Training of the H^2MCM model involves summing every 12 units of the *1-hr Prediction Model* and is described as follows:

1) TRAINING OF THE BASE 1-HR PREDICTION MODEL

The *Base 1-hr Prediction Model* is trained using the following inputs: a) Outliers, b) Initial Features, and c) Label_1. The output, *Predicted Feature of Base 1-hr Prediction Model*, is used as the input for subsequent *1-hr Prediction Models*.

2) TRAINING OF THE 1-HR PREDICTION MODEL_1

It uses the following inputs: a) Predicted Feature of Base 1-hr Prediction Model, b) Current Features, c) Outliers, d)

Label_1 (corresponding to the instance of the next hour from the current adjacent time), and e) heuristic rules.

After training, the *1-hr Prediction Model*$_1$ evaluates the errors using iterative heuristics and a feed-forward weight adjustment algorithm, *Delta-Rule*. This H^2MCM model predicts the APCs and MFs after 1 hour.

3) TRAINING OF THE 1-HR PREDICTION MODEL$_2$

For the second-hour prediction, *1-hr Prediction Model*$_2$' is trained using: a) the outputs of Predicted Feature of Base *1-hr Prediction Model* and 'Predicted Feature of *1-hr Prediction Model*$_1$', b) current features, c) outliers, d) Label_2 (corresponding to the 2-hour time interval from the current adjacent time), and e) heuristic rules.

The *1-hr Prediction Model*$_2$' computes errors and predicts the APCs and MFs after 2 hours. This process is repeated for n-hr Prediction Model_n' to predict features after 12 hours, and the results are summed to achieve APCs and MFs after 12 hours.

Algorithm 2 outlines our proposed H^2MCM model.

Algorithm 2 Proposed H^2MCM Model

Gather historical air quality data and relevant features from the constructed Big dataset
Pre-process the constructed Big dataset (e.g., handle missing values, scale features)
Define the H^2MCM model architecture
Split the big dataset into training, testing, and validation sets
Set the number of epochs and batch size for training
Train H^2MCM model:
    Initialize H^2MCM model with hyperparameters
    for each epoch **do**
        Feed training data to the H^2MCM model
        Update model weights using backpropagation
    end for
Make predictions:
    Input the constructed *Big Dataset*
    for each future time step **do**
        Use the trained H^2MCM model to predict air quality
        Apply transformer-based multi-chain hyper-heuristic rules using Algorithm 1
    end for
Evaluate the H^2MCM model:
    Compare predicted air quality with actual values on the test set
    Calculate performance metrics (e.g., PLCC, RMSE)
Output: H^2MCM model for future air quality prediction

By incorporating transformer-based methods, the H^2MCM model aims to leverage recent advancements to improve the accuracy and robustness of air quality predictions.

VI. EXPERIMENTAL RESULT

This section deals with the performance of our proposed H^2MCM model.

A. BIG DATASET

We utilized air quality and weather datasets obtained from Kaggle [4], [19]. The air quality dataset consists of hourly and daily concentration values for various APCs. Meanwhile, the weather dataset contains hourly and daily values for different meteorological factors (MFs), such as temperature (in Celsius), humidity, wind speed (in km/h), wind bearing (in degrees), visibility (in km), and cloud cover. The datasets were sourced from multiple sensors, comprising a total of 1,001,976 samples.

To create the required time-series dataset, we merged and pre-processed the air quality and weather datasets, followed by data normalization. Additionally, we integrated a noise dataset, generated by randomly sampling from the air quality and weather datasets. During training, we iteratively introduced 1,001,976 noise samples as outliers in each *1-hr Prediction Model*. Thus, every *1-hr Prediction Model* comprises 1,001,976 samples. As a result, the constructed dataset contains a total of 14,027,664 samples, which exhibit high heterogeneity, representing varying volumes, variety, and dynamics of the APCs.

B. SENSITIVITY ANALYSIS

In order to create a precise and reliable prediction model for the APCs, we construct a time-series big dataset by merging noise, air quality, and weather datasets, which demonstrate internal correlations. To assess the variability level of this combined dataset, we employ the *squared correlation coefficient* (R^2) metric, a standard tool for sensitivity analysis. The computation of the R^2 metric is as follows:

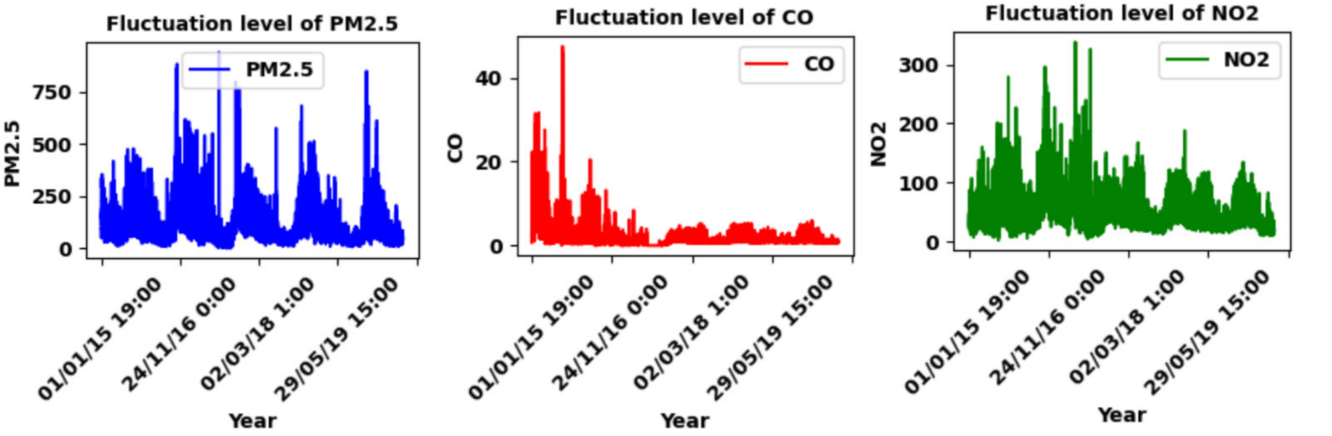
$$R^2 = 1 - \frac{\sum_{i=1}^{M}(ob_i - \hat{pd_i})^2}{\sum_{i=1}^{M}(ob_i - \bar{ob})^2} \tag{13}$$

Here, ob_i denotes the observed value, $\hat{pd_i}$ represents the predicted value, $\bar{ob}$ is the mean of the observed values, and M is the total number of data points. The symbol $\sum$ indicates the sum over all i data points.

In order to create an accurate air quality predictor, it is crucial to assess the present fluctuation levels of various participating air contaminants, as illustrated in Equation (13). Thus, a dependable predictor becomes indispensable to effectively handle these fluctuations and ensure precise forecasts of future air quality. Figure 2 depicts the current fluctuation levels of various participating air contaminants i.e., $PM_{2.5}$, CO, and NO_2.

The fluctuation levels of different participating air contaminants can be represented mathematically as follows: Let $C_{\text{contaminant}}(t)$ be the concentration of a specific air contaminant at time t. The fluctuation level of this contaminant can be calculated using the standard deviation $\sigma_{\text{contaminant}}$ and mean $\mu_{\text{contaminant}}$ as:

$$\text{Fluctuation level of contaminant:} \quad \text{Fluc}_{\text{contaminant}} = \frac{\sigma_{\text{contaminant}}}{\mu_{\text{contaminant}}} \tag{14}$$



(a) (b) (c)

FIGURE 2. (a) Current fluctuation level of $PM_{2.5}$. (b) Current fluctuation level of CO. (c) Current fluctuation level of NO_2.

TABLE 5. Best hyperparameter configuration.

Hyperparameter	Value
Loss	Validation MSE
Optimizer	Softmax
Metrics	RMSE, and PLCC
Batch size	68
Time step	1
Epochs	1000
Number LSTM layer	300
Learning Rate	0.211
Dropout	0.2

where:

$$\sigma_{\text{contaminant}} = \sqrt{\frac{1}{M}\sum_{tm=1}^{M}(C_{\text{contaminant}}(tm) - \mu_{\text{contaminant}})^2}$$

$$\mu_{\text{contaminant}} = \frac{1}{M}\sum_{tm=1}^{M} C_{\text{contaminant}}(tm)$$

Here, M represents the total number of data points (time steps (tm)) for the specific air contaminant.

C. EXPERIMENTAL SETUP

The experiments were executed on a server with an Intel i5 CPU and an NVIDIA Geforce RTX 3070 Ti GPU. Python 3.7 along with the required Anaconda environments served as the software environment for the experiments. To demonstrate the exceptional performance of our proposed H^2MCM model, we utilized the hyperparameter setup that yielded the best results, as indicated in Table 5.

D. EVALUATION METRICS

Our objective is to address the internal correlations among the participating APCs and MFs to achieve accurate predictions. To assess prediction accuracy, we have chosen the Pearson Linear Correlation Coefficient (PLCC) metric. PLCC is a real number ranging from '−1 to +1,' indicating the strength and direction of correlation between the APCs and MFs. A value of '+1' indicates a powerful positive correlation, while '−1' signifies a strong negative correlation.

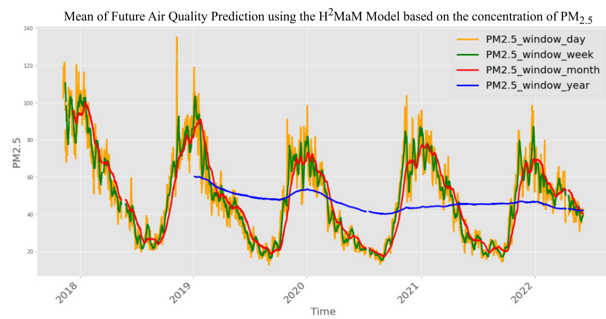


FIGURE 3. Mean of future air quality prediction using the H^2MCM model for 12 hours.

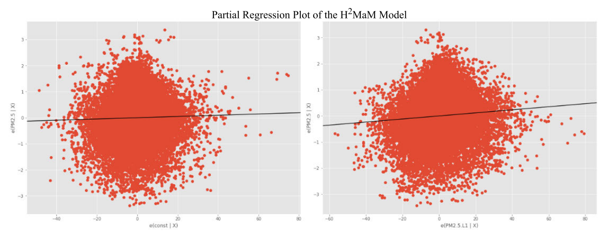


FIGURE 4. Partial regression of the H^2MCM model for 12 hours.

For evaluating prediction errors, we utilize the Root Mean Square Error (RMSE), which is a standard deviation with a range of '0 to 1'. A lower error is represented by '0', while a higher error is indicated by '1'. As a measure of prediction consistency, we employ RMSE as one of our evaluation metrics. Hence, a good prediction model should exhibit PLCC and RMSE values close to '1' and '0', respectively.

RMSE and PLCC can be computed as follows:

$$\text{RMSE} = \sqrt[e]{(1/M)\sum_{i=1}^{M}(pd_i - ob_i)^2} \quad (15)$$

$$\text{PLCC} = \sum_{i=1}^{M} \frac{\sum_{i=1}^{M}(pd_i - \overline{pd_i}) * (ob_i - \overline{ob_i})}{\sqrt[e]{\sum_{i=1}^{M}(pd_i - \overline{pd_i})^2 * \sum_{i=1}^{M}(ob_i - \overline{ob_i})^2}} \quad (16)$$

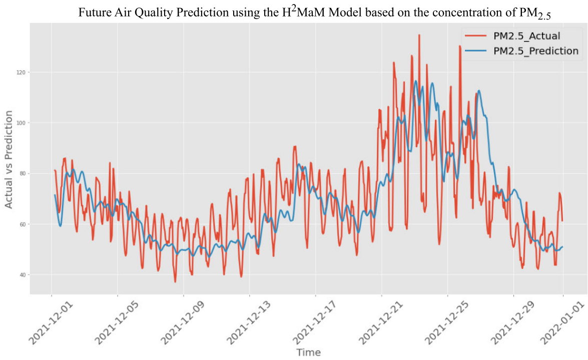


FIGURE 5. **Future air quality prediction using the H^2MCM model for 12 hours.**

Here, M represents the total number of samples. pd and ob refer to the predicted and observed values. $\overline{pd}$ and $\overline{ob}$ denote the mean of pd and ob, respectively.

The following models have been employed for comparison in this study: i) *Vlachogianni model* [9]: This model utilizes step-wise multiple linear regression for air quality prediction. ii) *RAQP model* [17]: The RAQP model applies recurrent support vector regression (SVR) for air quality prediction. iii) *SVR model* [12]: This model employs support vector regression for air quality prediction. iv) *MLP model* [13]: The MLP model is based on a multilayer perceptron for air quality prediction.

A comparison scheme was applied to all the regressors, utilizing the same features for evaluation.

E. EVALUATION OF THE H^2MCM MODEL

Performance evaluation of our proposed H^2MCM model is done by measuring the loss of the H^2MCM model through the mean square error (MSE) and mean absolute error (MAE) metrics. Figures 6(a) and 6(b) depict the MSE and MAE loss of the H^2MCM model, respectively.

F. PARSIMONY OF THE H^2MCM MODEL

To assess the parsimony of the H^2MCM Model, we used the *Occam's razor principle*. *Occam's razor principle* involves evaluating the model's complexity, simplicity, and adherence to the principle that the simplest explanation or model is often the best. Algorithm 3 describes the parsimony assessment of our proposed H^2MCM Model.

Table 6 presents the parsimony comparison among all used models with our proposed H^2MCM model.

G. PERFORMANCE EVALUATION

The big dataset was partitioned randomly into training, testing, and validation sets to effectively assess the performance of the *1-hr Prediction Model*. The training, testing, and validation sets contained 80%, 10%, and 10% of the data, respectively. The evaluation process was repeated 300 times, and the average PLCC and RMSE values for each APCs were computed from t=1 hr to t=12 hrs. The corresponding results are shown in tables 7 and 8.

Algorithm 3 Parsimony Assessment of the H^2MCM Model

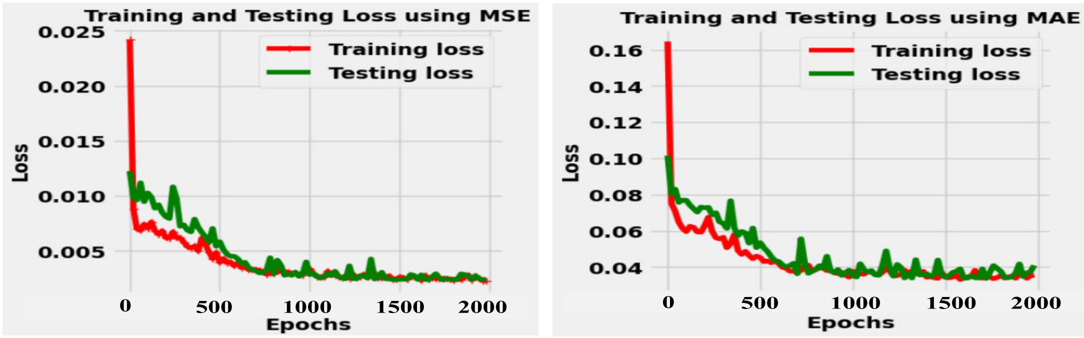
procedure ComputeParsimony(H^2MCM Model)
  $ComplexityScore(CS) \leftarrow 0$
  $InterpretabilityScore(IS) \leftarrow 0$
  $PerformanceScore(PS) \leftarrow 0$
  $CS \leftarrow EvaluateModelComplexity(H^2MCM\ Model)$
  $IS \leftarrow EvaluateInterpretability(H^2MCM\ Model)$
  $PS \leftarrow EvaluatePerformance(H^2MCM\ Model)$
  $TotalParsimonyScore \leftarrow CS + IS + PS$
  if *TotalParsimonyScore* is high **then**
    Output: The H^2MCM model is considered parsimonious according to Occam's razor principle.
  else
    Output: The H^2MCM model may need further simplification or improvement.
  end if
end procedure

Figure 3 displays the mean of future air quality predictions using the H^2MCM model, while Figure 4 depicts the partial regression of the H^2MCM model. Additionally, Figure 5 illustrates the future air quality predictions of the H^2MCM model.

Figure 9 clearly demonstrates the improved performance of our proposed H^2MCM model, accurately predicting future air pollution levels with PLCC and RMSE values close to 1 and 0, respectively. In contrast, the SVR, MLP, RAQP, and Vlachogianni models perform well only for shorter time intervals due to multi-chain hyper heuristic rules and the high correlation between APC and MF values. As a result, H^2MCM outperforms all the compared SVR, MLP, RAQP, and Vlachogianni models, as shown in Figures 7 and 8. We also consider the error caused by high internal correlations during each neighboring moment and find that our H^2MCM model achieves a 100% level of accuracy, making it one of the most reliable predictors for short-term air quality prediction. However, in reality, various uncontrollable natural factors like acid rain, cloudiness, humidity, and wind patterns prevent us from achieving 100% accurate forecasts.

Upon analyzing Figures 7 and 8, we observed that all models perform well for short-term air quality predictions due to the strong internal correlation between the participating APCs and MFs. However, as the time interval increases to medium- and long-term, the internal correlation weakens, leading to a decline in model performance. To address this limitation, H^2MCM incorporates multi-chain heuristic rules, which may result in some loss of information during prediction computation. To account for errors due to the accumulation and diffusion of information, we use standard metrics. Based on our findings, we can conclude that H^2MCM effectively handles the internal correlation between the participating APCs and MFs, outperforming other SVR, MLP, RAQP, and Vlachogianni models, and predicts future air quality with high accuracy.

Both ML- and DL-based regressors face error accumulation and diffusion issues due to noise/outlier features. However, our proposed H^2MCM model leverages advancements

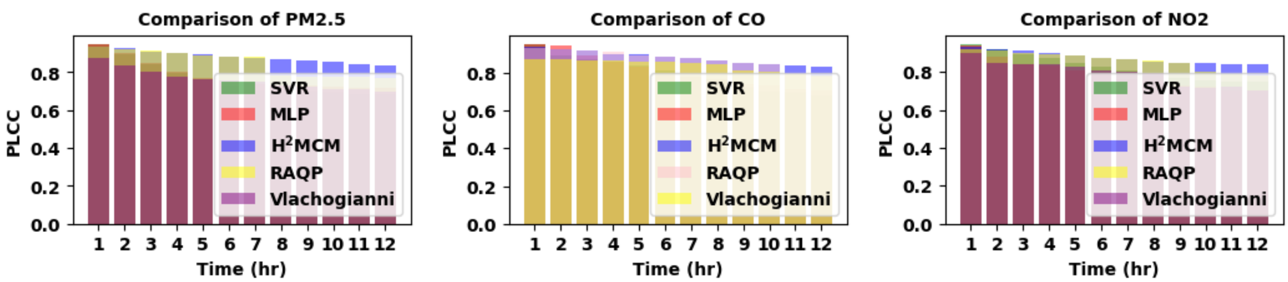


(a) (b)

FIGURE 6. (a) MSE of H^2MCM Model. (b) MAE of H^2MCM Model.

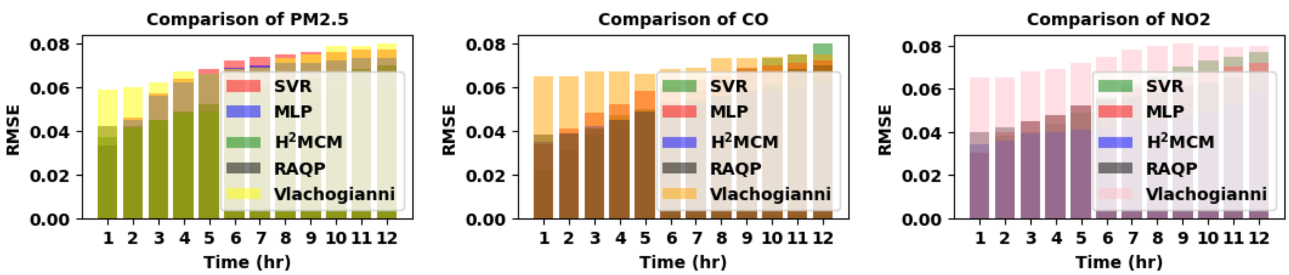
TABLE 6. Parsimony comparison among all models.

Model	Complexity Score	Interpretability Score	Performance Score	Total Parsimony Score	Parsimony Evaluation
SVR [12]	6	3	7	16	Parsimonious
MLP [13]	7	2	9	18	Not Parsimonious
RAQP [17]	4	5	6	15	Parsimonious
Vlachogianni [9]	6	4	7	17	Parsimonious
H^2MCM (Proposed)	**3**	**3**	**5**	**11**	**Parsimonious**



(a) (b) (c)

FIGURE 7. (a) Comparison of prediction accuracy of $PM_{2.5}$ from time step t=1 hr to t=12 hrs. (b) Comparison of Prediction Accuracy of CO from time step t=1 hr to t=12 hrs. (c) Comparison of Prediction Accuracy of NO_2 from time step t=1 hr to t=12 hrs.



(a) (b) (c)

FIGURE 8. (a) Comparison of prediction consistency of $PM_{2.5}$ from time step t=1 hr to t=12 hrs. (b) Comparison of Prediction Consistency of CO from time step t=1 hr to t=12 hrs. (c) Comparison of Prediction Consistency of NO_2 from time step t=1 hr to t=12 hrs.

in both the *1-hr Prediction Model* and deep techniques to overcome these issues and performs better than the SVR and MLP strategies. The results are validated in Figures 7 and 8, which demonstrate the superiority of our H^2MCM model over other SVR, MLP, RAQP, and Vlachogianni models. Nonetheless, due to the 1-*hr Prediction Model* and unavoidable natural factors like acid rain and volcanoes, the H^2MCM model may not perform well for certain time intervals.

Table 6 presents the compared parsimony results of five different models: SVR, MLP, RAQP, Vlachogianni, and H^2MCM models. The evaluation is based on three criteria: Complexity Score, Interpretability Score, and Performance Score, with a *Total Parsimony Score* being the sum of these three scores. *The Parsimony Evaluation* column provides a simple evaluation based on the total score, where a higher total score indicates less parsimony and a lower score indicates greater parsimony of models.

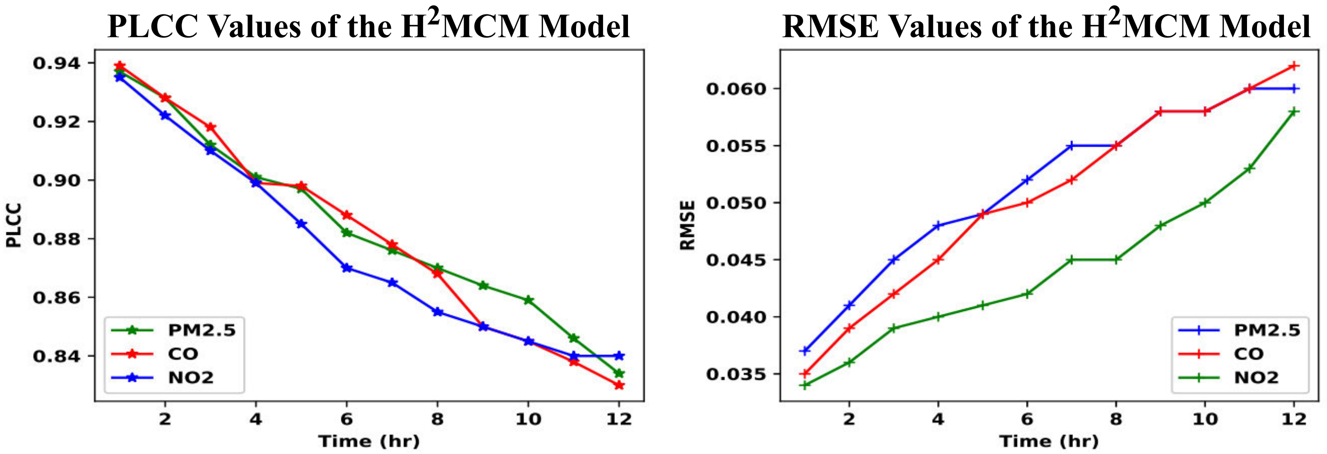


FIGURE 9. PLCC and RMSE values of the H^2MCM Model for 12 hours.

TABLE 7. PLCC comparison among ML-based SVR, DL-based MLP, RAQP, H^2MCM, and Vlachogianni models.

APCs	Method	t=1hr	t=2hr	t=3hr	t=4hr	t=5hr	t=6hr	t=7hr	t=8hr	t=9hr	t=10hr	t=11hr	t=12hr
$PM_{2.5}$	SVR [12]	0.949	0.896	0.841	0.795	0.761	0.721	0.707	0.699	0.697	0.693	0.687	0.685
	MLP [13]	0.949	0.899	0.847	0.803	0.773	0.743	0.735	0.731	0.732	0.726	0.719	0.715
	RAQP [17]	0.932	0.920	**0.915**	**0.902**	0.890	0.885	0.880	0.783	0.780	0.778	0.774	0.773
	Vlachogianni [9]	0.879	0.836	0.802	0.777	0.766	0.758	0.749	0.733	0.721	0.713	0.709	0.700
	H^2MCM(Proposed)	**0.957**	**0.928**	0.912	0.901	**0.897**	**0.892**	**0.886**	**0.870**	**0.864**	**0.859**	**0.846**	**0.834**
NO_2	SVR [12]	0.920	0.915	0.902	0.890	0.885	0.875	0.865	0.859	0.845	0.800	0.790	0.780
	MLP [13]	0.940	0.880	0.840	0.834	0.810	0.786	0.762	0.756	0.730	0.723	0.705	0.700
	RAQP [17]	0.986	0.961	**0.931**	**0.903**	0.886	0.869	0.843	0.830	0.800	0.780	0.750	0.709
	Vlachogianni [9]	0.898	0.850	0.842	0.837	0.825	0.805	0.799	0.793	0.721	0.713	0.719	0.705
	H^2MCM(Proposed)	**0.955**	**0.922**	0.905	0.890	**0.887**	**0.876**	**0.867**	**0.860**	**0.850**	**0.845**	**0.840**	**0.840**
CO	SVR [12]	0.95	0.89	0.871	0.856	0.837	0.810	0.780	0.764	0.723	0.698	0.690	0.680
	MLP [13]	0.952	0.948	0.890	0.860	0.830	0.800	0.780	0.763	0.740	0.733	0.715	0.708
	RAQP [17]	0.930	0.925	**0.915**	**0.900**	0.895	0.885	0.879	0.863	0.855	0.846	0.801	0.780
	Vlachogianni [9]	0.872	0.869	0.864	0.863	0.860	0.858	0.850	0.843	0.815	0.806	0.791	0.786
	H^2MCM(Proposed)	**0.979**	**0.958**	0.918	0.899	**0.898**	**0.888**	**0.880**	**0.868**	**0.859**	**0.849**	**0.838**	**0.830**

TABLE 8. RMSE comparison among direct ML-based SVR, direct DL-based MLP, RAQP, H^2MCM, and Vlachogianni models.

APCs	Method	t=1hr	t=2hr	t=3hr	t=4hr	t=5hr	t=6hr	t=7hr	t=8hr	t=9hr	t=10hr	t=11hr	t=12hr
$PM_{2.5}$	SVR [12]	0.033	0.046	0.057	0.064	0.068	0.072	0.074	0.075	0.076	0.076	0.077	0.077
	MLP [13]	0.033	0.045	0.056	0.062	0.066	0.069	0.07	0.071	0.071	0.072	0.073	0.073
	RAQP [17]	0.042	0.042	**0.048**	**0.048**	0.052	0.059	0.061	0.062	0.064	0.066	0.068	0.070
	Vlachogianni [9]	0.059	0.060	0.062	0.067	0.066	0.068	0.069	0.073	0.075	0.079	0.079	0.080
	H^2MCM(Proposed)	**0.030**	**0.041**	0.049	0.049	**0.049**	**0.052**	**0.055**	**0.055**	**0.058**	**0.058**	**0.060**	**0.060**
NO_2	SVR [12]	0.030	0.038	0.040	0.044	0.049	0.056	0.062	0.066	0.070	0.073	0.075	0.077
	MLP [13]	0.030	0.040	0.045	0.048	0.052	0.055	0.060	0.063	0.065	0.068	0.070	0.072
	RAQP [17]	0.032	0.035	**0.039**	**0.039**	0.045	0.048	0.052	0.058	0.061	0.063	0.065	0.068
	Vlachogianni [9]	0.065	0.065	0.068	0.069	0.072	0.075	0.078	0.080	0.081	0.080	0.079	0.080
	H^2MCM(Proposed)	**0.029**	**0.036**	0.039	0.040	**0.041**	**0.042**	**0.045**	**0.045**	**0.048**	**0.050**	**0.053**	**0.058**
CO	SVR [12]	0.022	0.031	0.038	0.047	0.050	0.053	0.058	0.063	0.068	0.073	0.075	0.080
	MLP [13]	0.034	0.041	0.048	0.052	0.058	0.060	0.063	0.065	0.069	0.070	0.071	0.072
	RAQP [17]	0.038	0.040	**0.041**	**0.044**	0.050	0.052	0.054	0.056	0.059	0.061	0.068	0.070
	Vlachogianni [9]	0.065	0.065	0.067	0.067	0.066	0.068	0.069	0.073	0.073	0.074	0.075	0.075
	H^2MCM(Proposed)	**0.035**	**0.039**	0.042	0.045	**0.049**	**0.050**	**0.052**	**0.055**	**0.058**	**0.058**	**0.060**	**0.062**

SVR, RAQP, and H^2MCM models are considered parsimonious because they have relatively lower total parsimony scores, indicating that they strike a better balance between complexity, interpretability, and performance. On the other hand, the MLP model is not considered parsimonious because it has the highest total parsimony score, indicating that it is relatively more complex, less interpretable, or has lower performance according to the chosen criteria. Vlachogianni model is also considered parsimonious, with a score similar to SVR, RAQP, and H^2MCM models. The H^2MCM model provides greater parsimony among all compares models. Therefore, from Table 6, we can say that the H^2MCM model is superior among all compared models.

H. SO-WHAT ASPECT

Our research introduces the Hyper Heuristic Multi-Chain Model (H^2MCM), a novel and highly advanced approach for air quality forecasting. This is not just an academic exercise; it has real-world implications and significance. The *so-what* aspect of our research can be framed as follows:

1) **Public Health and Environmental Protection:** H^2MCM's accurate predictions have the potential to significantly improve public health by helping individuals avoid exposure to harmful air pollutants. By extension, this can lead to a reduction in healthcare costs and an enhancement in the quality of life for the population.
2) **Mitigating Climate Change:** Improved air quality forecasting contributes to our understanding of the relationship between air pollutants and climate change. The model can support initiatives aimed at reducing greenhouse gas emissions and mitigating climate change.
3) **Urban Planning and Infrastructure Development:** City planners can use precise air quality forecasts to optimize the placement of urban infrastructure, minimizing residents' exposure to pollution and improving the overall quality of life in cities.
4) **Emergency Response:** H^2MCM's accuracy is crucial for emergency response during environmental disasters. It can provide timely information for evacuations and resource allocation, potentially saving lives.
5) **Economic Benefits:** Accurate air quality predictions have economic implications. They can lead to cost savings by reducing healthcare expenses, increasing worker productivity, and minimizing damage to crops and buildings, resulting in economic benefits for both individuals and businesses.
6) **Influencing Policy and Regulation:** Policymakers can utilize the findings from H^2MCM to develop or modify air quality regulations and policies, ultimately improving air quality standards and protecting the environment.
7) **Future Research and Innovation:** This research paves the way for future studies and innovations in air quality prediction. It can inspire the development of new technologies, sensors, and data sources to advance our understanding of air quality and environmental sustainability.
8) **Local and Global Impact:** The research is significant both at the local and global levels. It addresses specific air quality challenges in the region where it is applied, but it also contributes to the global effort to address air quality issues and environmental sustainability.

The *so-what* aspect of this research goes beyond the technical details of the model. It highlights the practical significance and real-world impact of accurate air quality forecasting, emphasizing the benefits to public health, the environment, the economy, and decision-making at various levels of society.

VII. DISCUSSION

In this study, we introduce a novel per hour heuristic multi-chain mechanism, applicable in industrial scenarios, such as acoustic compression, to predict the concentration of various air contaminants. Our methodology incorporates multiple features, including adjacent time intervals, temperature, relative humidity, wind speed, pressure, CO, NO_2, and $PM_{2.5}$ to achieve efficient and precise forecasts. Each feature is computed independently for its immediate succeeding time interval, and the overall air quality score is determined based on the anticipated APCs after several hours. By employing this multi-chain process, we can predict APCs and air quality index (AQI) several hours in advance. Our assessment results validate the accuracy and precision of the H^2MCM model compared to other models. However, the H^2MCM model might exhibit reduced performance due to the utilized compression technique and the prediction gap between the 1-hour prediction models and perfect 100% performance. To bridge this gap, we propose an ensemble learning method that integrates various learning algorithms such as XGBoost, LSTM, GRU, and BGRU, known for their robust memorization and performance-boosting capabilities. Through this ensemble approach, we enhance the predictive performance of our H^2MCM model.

VIII. CONCLUSION AND FUTURE WORK

Our study addresses the urgent concern of air quality prediction, which requires immediate attention. To tackle this challenge, we introduce an innovative heuristic hourly-based multi-chain strategy. Also, for an efficient model selection and decision-making process, this study uses the Akaike Information Criterion (AIC), Schwarz Bayesian Information Criterion (SBIC), Hannan-Quinn Information Criterion (HQIC), and corrected AIC (AICc) tools due to their ability to balance model fit and complexity. These criteria provide a quantitative measure to compare different models based on their goodness of fit while penalizing the number of parameters in the model. By considering both the fit and complexity of the models, we can select the best model to balance explaining the data well and avoiding overfitting. In essence, AIC, SBIC, HQIC, and AICc help researchers identify the most parsimonious model that adequately explains the observed data, thereby aiding in selecting models likely to generalize well to new data. The effectiveness of our H^2MCM model in predicting air quality is well-established, as evident from its impressive PLCC and RMSE metrics. Through experimentation, we have demonstrated that the H^2MCM model surpasses SVR, MLP, RAQP, and Vlachogianni models by 75%, 60%, 38%, and 70%, respectively.

However, we acknowledge that while the H^2MCM model exhibits exceptional performance, it may not always be the optimal choice for every scenario. Thus, our future plans involve delving deeper into the non-linear properties of APCs and MFs using real datasets from urban areas. Furthermore, we are enthusiastic about validating our 1-hr prediction model using IoT sensors within extensive real-world datasets. This endeavor will enable us to iteratively refine and augment the H^2MCM model's functionalities, making it more adept for real-world applications. Also, we want to incorporate the insights from "Long-range dependence and heavy tail characteristics for remaining useful life prediction in rolling bearing degradation" into the H^2MCM model by following these steps:

1) *Feature Integration:* Extract features related to long-range dependence and heavy tail characteristics from the used dataset. These features could include statistical measures capturing the distributional properties of the degradation process, such as skewness, kurtosis, and autocorrelation.
2) *Model Adaptation:*
   - To adapt the H^2MCM architecture to incorporate features related to long-range dependence and heavy tail characteristics, adjustments to the input layer and heuristic selection mechanism are necessary as described below:
     - *Adjusting the Input Layer:* Expand the input layer of the H^2MCM model from 12 to 24 for accommodating the new features derived from long-range dependence and heavy tail characteristics.
     - *Integrating Features into Heuristic Selection:* We modify the heuristic selection mechanism based on *feature gating* to incorporate information from the newly added features to dynamically adjust its heuristic choices based on the values of these features.
3) *Training and Validation:*
   - Reconfigure the training process by changing the hyperparameter setup to include the new features and ensure that the H^2MCM model learns to utilize them for prediction effectively.
   - Validate the adapted H^2MCM using 10-fold cross-validation techniques, considering its performance on historical data and its ability to generalize to unseen data.
4) *Evaluation and Comparison:*
   - Evaluate the performance of the adapted H^2MCM model against the original version and other relevant models based on the PLCC and RMSE metrics.
   - Compare the predictive accuracy, robustness, and efficiency of the adapted H^2MCM model with and without incorporating long-range dependence and heavy tail characteristics.

By following these steps, we can effectively incorporate the insights from the referenced study into the H^2MCM model and potentially enhance its predictive capabilities for remaining useful life prediction in rolling bearing degradation.

APPENDIX A
EXPLANATIONS OF THE USED ABBREVIATIONS

Table 9 describes concise explanations for all the abbreviations utilized in this study.

TABLE 9. Abbreviations and explanations.

Abbreviation	Explanation
H^2MCM	Combining the alphabets of Hyper Heuristic Multi-Chain Model
MFs	Meteorological Factors
APCs	Air Pollutants' Concentrations
1-hr Prediction Model	1-hour Prediction Model
AIC	Akaike Information Criterion is a tool for comparing and selecting the most appropriate model.
SBIC	Schwarz Bayesian Information Criterion is a tool for comparing and selecting the most appropriate model.
HQIC	Hannan-Quinn Information Criterion is a tool for comparing and selecting the most appropriate model.
AICc	corrected Akaike Information Criterion is a tool for comparing and selecting the most appropriate model.
SVR	Support Vector Regression is used here as an ML-based prediction approach.
MLP	Multi-Layer Perceptron is used here as a DL-based prediction approach.
RAQP	Recurrent Air Quality Predictor is used here as a compared model.
LSTM	Long short-term memory is the base of the proposed H^2MCM Model.
PLCC	PLCC stands for the Pearson Linear Correlation Coefficient. High PLCC value indicates a powerful positive correlation, while lower PLCC value indicates a negative correlation.
RMSE	RMSE stands for the Root Mean Square Error. Lower RMSE value indicates higher prediction consistency and vise-versa.
MSE	MSE is stands for the Mean Squared Error. MSE is used here as a risk function that measures the square errors of the H^2MCM model.
MAE	MAE is stands for the Mean Absolute Error. MAE is used here as a risk function that measures the absolute errors of the H^2MCM model.
R^2	R^2 stands for the Squared Correlation Coefficient metric. R^2 is used here as a standard tool for preforming the sensitivity analysis of the H^2MCM model.

REFERENCES

[1] Word Health Organization. *9 Out of 10 People Worldwide Breathe Polluted Air, But More Countries Are Taking Action*. Accessed: Aug. 25, 2023. [Online]. Available: https://www.who.int/health-topics/air-pollution#tab=tab_1

[2] K. Gu, H. Liu, J. Liu, X. Yu, T. Shi, and J. Qiao, "Air pollution prediction in mass rallies with a new temporally-weighted sample-based multitask learner," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–15, 2022.

[3] J. Saiohai, S. Bualert, T. Thongyen, K. Duangmal, P. Choomanee, and W. W. Szymanski, "Statistical $PM_{2.5}$ prediction in an urban area using vertical meteorological factors," *Atmosphere*, vol. 14, no. 3, p. 589, Mar. 2023.

[4] *Weather Dataset*. Accessed: Jul. 15, 2023. [Online]. Available: https://www.kaggle.com/datasets/muthuj7/weather-dataset

[5] Y. Yang, G. Mei, and S. Izzo, "Revealing influence of meteorological conditions on air quality prediction using explainable deep learning," *IEEE Access*, vol. 10, pp. 50755–50773, 2022.
[6] Y. Zhou, S. De, G. Ewa, C. Perera, and K. Moessner, "Data-driven air quality characterization for urban environments: A case study," *IEEE Access*, vol. 6, pp. 77996–78006, 2018.
[7] W. Tong, L. Li, X. Zhou, A. Hamilton, and K. Zhang, "Deep learning $PM_{2.5}$ concentrations with bidirectional LSTM RNN," *Air Quality Atmos. Health*, vol. 12, pp. 1–13, Apr. 2019.
[8] D. Voukantsis, K. Karatzas, J. Kukkonen, T. Räsänen, A. Karppinen, and M. Kolehmainen, "Intercomparison of air quality data using principal component analysis, and forecasting of PM_{10} and $PM_{2.5}$ concentrations using artificial neural networks, in Thessaloniki and Helsinki," *Sci. Total Environ.*, vol. 409, no. 7, pp. 1266–1276, Mar. 2011.
[9] A. Vlachogianni, P. Kassomenos, A. Karppinen, S. Karakitsios, and J. Kukkonen, "Evaluation of a multiple regression model for the forecasting of the concentrations of NO_x and PM_{10} in Athens and Helsinki," *Sci. Total Environ.*, vol. 409, no. 8, pp. 1559–1571, Mar. 2011.
[10] Y. Wang, Q. Ying, J. Hu, and H. Zhang, "Spatial and temporal variations of six criteria air pollutants in 31 provincial capital cities in China during 2013–2014," *Environ. Int.*, vol. 73, pp. 413–422, Dec. 2014.
[11] B. S. Freeman, G. Taylor, B. Gharabaghi, and J. Thé, "Forecasting air quality time series using deep learning," *J. Air Waste Manage. Assoc.*, vol. 68, no. 8, pp. 866–886, Aug. 2018.
[12] Y. Yusof and I. S. Maijamaa, "Air quality index prediction using support vector regression based on African buffalo optimization," in *Proc. Int. Conf. Comput. Inform.* Singapore: Springer, 2023, pp. 3–14.
[13] E. Solgi and M. P. Mehr, "Predicting and modeling of daily concentration of particulate matter ($PM_{2.5}$ & PM_{10}) in Hamadan winter with multilayer perceptron neural network," *Environ. Res.*, vol. 13, pp. 99–114, Feb. 2023.
[14] N. Ye, L. Cao, L. Yang, Z. Zhang, Z. Fang, Q. Gu, and G.-Z. Yang, "Improving the robustness of analog deep neural networks through a Bayes-optimized noise injection approach," *Commun. Eng.*, vol. 2, no. 1, p. 25, May 2023.
[15] H. Q. Bang, H. D. Nguyen, K. Vu, and T. T. Hien, "Photochemical smog modelling using the air pollution chemical transport model (TAPM-CTM) in Ho Chi Minh city, Vietnam," *Environ. Model. Assessment*, vol. 24, no. 3, pp. 295–310, Jun. 2019.
[16] H. Wang, "Air pollution and meteorological data in Beijing 2017–2018," Harvard Dataverse, V1, Mar. 2019, doi: 10.7910/DVN/USXCAK.
[17] K. Gu, J. Qiao, and W. Lin, "Recurrent air quality predictor based on meteorology-and pollution-related factors," *IEEE Trans. Ind. Informat.*, vol. 14, no. 9, pp. 3946–3955, Sep. 2018.
[18] K. B. Shaban, A. Kadri, and E. Rezk, "Urban air pollution monitoring system with forecasting models," *IEEE Sensors J.*, vol. 16, no. 8, pp. 2598–2606, Apr. 2016.
[19] *Air Quality Meteorological Data*. Accessed: Aug. 15, 2023. [Online]. Available: https://www.kaggle.com/rohanrao/air-quality-data-in-india
[20] B. T. Ong, K. Sugiura, and K. Zettsu, "Dynamically pre-trained deep recurrent neural networks using environmental monitoring data for predicting $PM_{2.5}$," *Neural Comput. Appl.*, vol. 27, no. 6, pp. 1553–1566, Aug. 2016.
[21] Y. Han, J. C. K. Lam, V. O. K. Li, and Q. Zhang, "A domain-specific Bayesian deep-learning approach for air pollution forecast," *IEEE Trans. Big Data*, vol. 8, no. 4, pp. 1034–1046, Aug. 2022.
[22] X. Yi, Z. Duan, R. Li, J. Zhang, T. Li, and Y. Zheng, "Predicting fine-grained air quality based on deep neural networks," *IEEE Trans. Big Data*, vol. 8, no. 5, pp. 1326–1339, Oct. 2022.
[23] J. Y. Zhu, C. Sun, and V. O. K. Li, "An extended spatio-temporal Granger causality model for air quality estimation with heterogeneous urban big data," *IEEE Trans. Big Data*, vol. 3, no. 3, pp. 307–319, Sep. 2017.
[24] W. C. Leong, R. O. Kelani, and Z. Ahmad, "Prediction of air pollution index (API) using support vector machine (SVM)," *J. Environ. Chem. Eng.*, vol. 8, no. 3, Jun. 2020, Art. no. 103208.
[25] Z. Luo, J. Huang, K. Hu, X. Li, and P. Zhang, "AccuAir: Winning solution to air quality prediction for KDD cup 2018," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Jul. 2019, pp. 1842–1850.
[26] X. Yi, J. Zhang, Z. Wang, T. Li, and Y. Zheng, "Deep distributed fusion network for air quality prediction," in *Proc. 24th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Jul. 2018, pp. 965–973.
[27] B. Liu, S. Yan, J. Li, G. Qu, Y. Li, J. Lang, and R. Gu, "An attention-based air quality forecasting method," in *Proc. 17th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA)*, Dec. 2018, pp. 728–733.
[28] W. Cheng, Y. Shen, Y. Zhu, and L. Huang, "A neural attention model for urban air quality inference: Learning the weights of monitoring stations," in *Proc. AAAI Conf. Artif. Intell.*, Apr. 2018, pp. 2151–2158.
[29] Y. Qi, Q. Li, H. Karimian, and D. Liu, "A hybrid model for spatiotemporal forecasting of $PM_{2.5}$ based on graph convolutional neural network and long short-term memory," *Sci. Total Environ.*, vol. 664, pp. 1–10, May 2019.
[30] C.-J. Huang and P.-H. Kuo, "A deep CNN-LSTM model for particulate matter ($PM_{2.5}$) forecasting in smart cities," *Sensors*, vol. 18, no. 7, p. 2220, Jul. 2018.
[31] P.-W. Chiang and S.-J. Horng, "Hybrid time-series framework for daily-based $PM_{2.5}$ forecasting," *IEEE Access*, vol. 9, pp. 104162–104176, 2021.
[32] Z. Qi, T. Wang, G. Song, W. Hu, X. Li, and Z. Zhang, "Deep air learning: Interpolation, prediction, and feature analysis of fine-grained air quality," *IEEE Trans. Knowl. Data Eng.*, vol. 30, no. 12, pp. 2285–2297, Dec. 2018.
[33] B. Liu, S. Yan, J. Li, Y. Li, J. Lang, and G. Qu, "A spatiotemporal recurrent neural network for prediction of atmospheric $PM_{2.5}$: A case study of Beijing," *IEEE Trans. Computat. Social Syst.*, vol. 8, no. 3, pp. 578–588, Jun. 2021.
[34] Q. Guo, Z. He, and Z. Wang, "Simulating daily $PM_{2.5}$ concentrations using wavelet analysis and artificial neural network with remote sensing and surface observation data," *Chemosphere*, vol. 340, Nov. 2023, Art. no. 139886.
[35] N. S. Gupta, Y. Mohta, K. Heda, R. Armaan, B. Valarmathi, and G. Arulkumaran, "Prediction of air quality index using machine learning techniques: A comparative analysis," *J. Environ. Public Health*, vol. 2023, pp. 1–26, Jan. 2023.
[36] P. Li, S. Wang, H. Ji, Y. Zhan, and H. Li, "Air quality index prediction based on an adaptive dynamic particle swarm optimized bidirectional gated recurrent neural network–China region," *Adv. Theory Simulations*, vol. 4, no. 12, Dec. 2021, Art. no. 2100220.
[37] J. A. Moscoso-López, J. González-Enrique, D. Urda, J. J. Ruiz-Aguilar, and I. J. Turias, "Hourly pollutants forecasting using a deep learning approach to obtain the AQI," *Log. J. IGPL*, vol. 31, no. 4, pp. 722–738, Jul. 2023.
[38] M. T. Udristioiu, Y. E. Mghouchi, and H. Yildizhan, "Prediction, modelling, and forecasting of PM and AQI using hybrid machine learning," *J. Cleaner Prod.*, vol. 421, Oct. 2023, Art. no. 138496.
[39] Y. Huang, J. J.-C. Ying, and V. S. Tseng, "Spatio-attention embedded recurrent neural network for air quality prediction," *Knowl.-Based Syst.*, vol. 233, Dec. 2021, Art. no. 107416.
[40] S. D. Yang, Z. A. Ali, and B. M. Wong, "FLUID-GPT (fast learning to understand and investigate dynamics with a generative pre-trained transformer): Efficient predictions of particle trajectories and erosion," *Ind. Eng. Chem. Res.*, vol. 62, no. 37, pp. 15278–15289, Sep. 2023.
[41] N. Geneva and N. Zabaras, "Transformers for modeling physical systems," *Neural Netw.*, vol. 146, pp. 272–289, Feb. 2022.

KALYAN CHATTERJEE (Member, IEEE) received the B.Tech. and M.Tech. degrees in computer science and engineering from State Government University, West Bengal, India. He is currently an Assistant Professor with the Department of Computer Science and Engineering, Nalla Malla Reddy Engineering College, Hyderabad, Telangana, India. He has more than seven years of teaching experience with B.Tech. colleges in India. He has three patents (Intellectual Property India). He has published journal and conference papers in esteemed international journals and conferences. His current research interests include air pollution, climate, the IoT, big data, and healthcare. He is an IEEE, ACM, and CSI Professional Body Member. He is also a Reviewer of repute international journals, such as MDPI.

SAMLA SURAJ KUMAR (Student Member, IEEE) is currently pursuing the B.Tech. degree with the Department of Artificial Intelligence and Machine Learning, NMREC, Hyderabad. He has two patents (Intellectual Property India), more than five conference cum book series, and two book chapters in international publications. His current research interests include climate science, artificial intelligence, and machine learning.

RAMAGIRI PRAVEEN KUMAR received the B.Tech. degree from the Department of Computer Science and Engineering, Mother Teresa College of Engineering and Technology, in 2012, the M.Tech. degree in nanotechnology from the Aurora's Scientific Technological and Research Institute, in 2014, and the Ph.D. degree from the Department of Nanoscience and Technology, The Central University of Jharkhand, in 2019. He is currently an Assistant Professor with the Department of Computer Science and Engineering, Nalla Malla Reddy Engineering College, Hyderabad. His research interests include to understand the physical phenomenon involved in different nano structures through employing the sputter deposition of materials on various kinds of substrate and examining their properties with indentation technique. He is involved in developing computational models on MATLAB.

ANJAN BANDYOPADHYAY received the M.Tech. degree in information security from the Department of Information Technology, NIT, Durgapur, West Bengal, India, and the Ph.D. degree from NIT. He is currently an Assistant Professor with the School of Computer Engineering, Kalinga Institute of Industrial Technology, Bhubaneswar, Odisha, India. He is broadly interested in algorithmic game theory (mechanism design). He has published many conference and journal papers in esteemed conferences and journals. His current research interests include cloud computing, fog computing, metaverse, the IoT, healthcare, and crowd sourcing. He completed the Vishveshwarya Ph.D. Fellowship under MHRD. He bags a number of best paper awards in many conferences, such as 3PGCIC.

SUJATA SWAIN received the B.Tech. degree in computer science and engineering from BPUT University, India, and the M.Tech. and Ph.D. degrees in computer science and engineering from Indian Institute of Technology Roorkee, India. She is currently an Assistant Professor with the School of Computer Engineering, Kalinga Institute of Industrial Technology, Deemed to be University, Bhubaneswar. Her research interests include service oriented computing, metaverse, medical imaging, and health-care systems.

SAURAV MALLIK (Member, IEEE) received the Ph.D. degree from the Department of Computer Science and Engineering, Jadavpur University, Kolkata, India, in 2017. His Ph.D. works were conducted with the Machine Intelligence Unit, Indian Statistical Institute, Kolkata. He was a Postdoctoral Fellow with Harvard T. H. Chan School of Public Health, The University of Texas Health Science Center at Houston, and the University of Miami Miller School of Medicine, Miami, FL, USA. He is currently a Research Scientist with The University of Arizona, USA. He has co-authored more than 125 research articles with a Google H-index of 20. His research interests include computational biology, bioinformatics, data mining, bio-statistics, machine learning, and pattern recognition. He was a recipient of the Research Associate of the Council of Scientific and Industrial Research, MHRD, Government of India, in 2017; and the Emerging Researcher in Bioinformatics Award from the Bioclues BIRD Award Steering Committee, India, in 2020. He is an editor of many journals.

AMAL AL-RASHEED received the Ph.D. degree in information systems from King Saud University, in 2017. She is currently an Associate Professor with the Department of Information Systems, College of Computer and Information Sciences, Princess Nourah Bint Abdulrahman University (PNU), Riyadh, Saudi Arabia. She has been involved in many projects related to learning technologies, cyber security, and virtual reality. Her contributions in research projects in academia led to the publication of papers in many journals and conferences. Her research interests include education, knowledge management, data mining, data analytics, cyber security, and natural language processing. In 2017, she was awarded the Research Excellence Award by PNU, for her publications during the Ph.D.

MOHAMED ABBAS received the B.Sc. degree in electronics engineering from College of Engineering, Mansoura University, Egypt, in 1998, and the M.Sc. and Ph.D. degrees in computer engineering from Mansoura University in 2002 and 2008, respectively. He worked as an Assistant Professor at the Department of Communications and Computer Engineering, College of Engineering, Delta University. He is currently working as a Full Professor with the Department of Electrical Engineering, King Khaled University, Abha, KSA. His areas of interest are intelligent systems, medical informatics, nanotechnology, and bioinformatics.

BEN OTHMAN SOUFIENE received the M.S. degree from Monastir University, in 2012, and the Ph.D. degree in computer science from Manouba University, in 2016, for his dissertation on "secure data aggregation in wireless sensor networks. His research interests include the Internet of Medical Things, wireless body sensor networks, wireless networks, artificial intelligence, machine learning, and big data.