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

Dr.AB.Hajira Be[1], G.Rithik Krishnan[2]

[1] Associate Professor
Department of Computer Applications
Karpaga Vinayaga College of Engineering and Technology
Maduranthagam TK
[2]PG Student
Department of Computer Applications
Karpaga Vinayaga College of Engineering and Technology
**Corresponding Author: Rithik Krishnan G Email: rithikcbcs@gmail.com*

---***---

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 (H2MCM) 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), H2MCM accurately predicts forthcoming air pollutants (APs) concentrations such as particulate matter with diameter 2.5 μm (PM2.5), carbon monoxide (CO), and nitrogen dioxide (NO2). Additionally, it accounts for spatiotemporal correlations between these APs and MFs, which significantly influence the air quality prediction for the next immediate time interval. H2MCM 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 H2MCM's superiority over various models, including the Support Vector Regressor (SVR), Multi-Layer Perceptron (MLP), Recurrent Air Quality Predictor (RAQP), and Valchogianni models. H2MCM 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. This research introduces a hybrid deep learning model for precise air quality prediction, crucial for effective environmental monitoring. The model integrates Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTMs) to capture both spatial and temporal dependencies within air pollution data, utilizing a comprehensive dataset from Kaggle containing PM2.5, CO, NO2, and SO2 measurements. Initially, data preprocessing addresses missing values and normalizes features. Subsequently, a CNN extracts spatial patterns, identifying relationships between pollutants, while an LSTM analyzes temporal sequences, capturing air quality evolution.

Key Words: AI-powered digital footprint, Data privacy, Cyber security, risks Identity protection.

1. INTRODUCTION

This urban development and industrialization have led to a severe air pollution problem in urban areas, posing 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. 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. The conventional chemical transport model (CTM) 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 have presented deep learning-based non- linear air pollution models, while others have successfully applied machine learning-based repressors to predict future air quality. 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) technique performs well for short time intervals due to strong correlations among air pollution components (APCs) and meteorological factors (MFs). However, this correlation weakens as the time intervals increase, rendering SVR 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 H2MCM model. Each *1-hr Prediction Model* is needed to create the *chain* architecture of the H2MCM 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 H2MCM model with the LSTM regressor to capture non-linear relation- ships among APCs and MFs

To construct the big dataset, random sampling techniques are used, and noise/outlier sets are added to the original dataset as described. 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

2. RELATED WORKS

The H2MCM 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 H2MCM 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.

1. Limitations of Integrating Transformer Models with H2MCM

- Computational Complexity:

  a. Transformers: Transformer models, such as those discussed 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. H2MCM (LSTM-based): LSTM models are generally more efficient at handling long sequences because they process sequences sequentially and maintain a manageable computational complexity.

- 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. H2MCM (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.

- Interpretability:

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

  b. H2MCM (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.

- 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 embedding.

  b. H2MCM (LSTM-based): LSTM models are inherently designed to capture temporal dependencies, making them well-suited for time series data like APCs and MFs.

2. Transformer-Based Multi-Chain Hyper-Heuristic Rules

The optimization process aims to find the best combination of transformer hyperparameters that yield the optimal performance for the proposed H2MCM model. The transformer-based multichain hyper-heuristic lever- ages 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 H2MCM model can enhance the accuracy and consistency of air quality prediction. The proposed transformer-based multi-chain hyper-heuristic rules to develop the H2MCM model.

3. Training of the H^2MCM Model

Training of the H2MCM model involves summing every 12 units of the *1-hr Prediction Model* and is described as follows:

- 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*.
- 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.
- Raining 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.

3. PERFORMANCE

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

1. Big Dataset

We utilized air quality and weather datasets obtained from Kaggle. 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.

2. 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.

$$R^2 = 1 - \frac{\sum_{i=1}^{M}(O_{bi} - \hat{P}di)^2}{\sum_{i=1}^{M}(ob_i - \overline{ob})^2}$$

Here, ob_i denotes the observed value, $p\,\hat{}\,d_i$ represents the predicted value, $o\,\bar{}\,b$ is the mean of the observed values, and M is the total number of data points. The symbol 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. Thus, a dependable predictor becomes indispensable to effectively handle these fluctuations and ensure precise forecasts of future air quality. 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 μcontaminant.

3. 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.

4. 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.

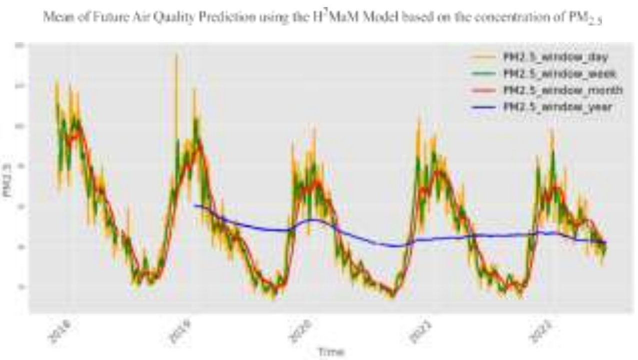


Figure 1. Mean of future air quality prediction using the H2MCM model for 12 hours

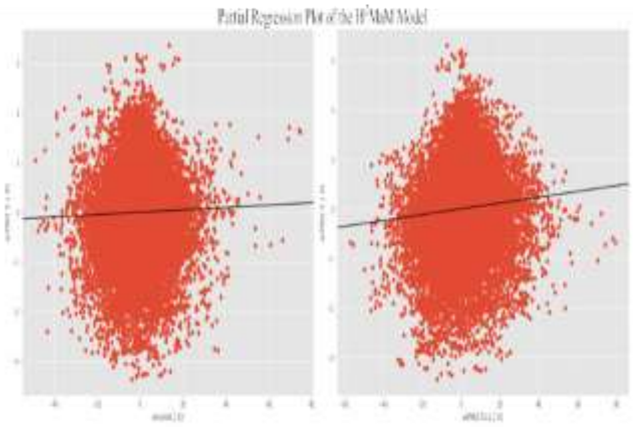


Figure 2. Partial regression of the H2MCM model for 12 hours

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.

5. Evaluation of the H^{2}MCM Model

Performance evaluation of our proposed H^{2}MCM model is done by measuring the loss of the H^{2}MCM model through the mean square error (MSE) and mean absolute error (MAE) metrics

6. Parsimony of the H^{2}MCM Model

To assess the parsimony of the H^{2}MCM 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.

7. 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.

8. SO-WHAT Aspect

Our research introduces the Hyper Heuristic Multi-Chain Model (H2MCM), 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:

- **Public Health and Environmental Protection:** H^{2}MCM’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.
- **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.
- **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.
- **Emergency Response:** H^{2}MCM’s accuracy is crucial for emergency response during environmental disasters. It can provide timely information for evacuations and resource allocation, potentially saving lives.
- **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.
- **Influencing Policy and Regulation:** Policymakers can utilize the findings from H^{2}MCM to develop or modify air quality regulations and policies, ultimately improving air quality standards and protecting the environment.
- **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.
- **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.

4. PROPOSED SYSTEM

To address the limitations of existing systems, this project proposes a hybrid deep learning model that combines the strengths of CNNs and LSTMs for accurate air quality prediction. The proposed system leverages the spatial feature extraction capabilities of CNNs and the temporal modeling prowess of LSTMs to capture the complex relationships and dynamic evolution of air pollutants.

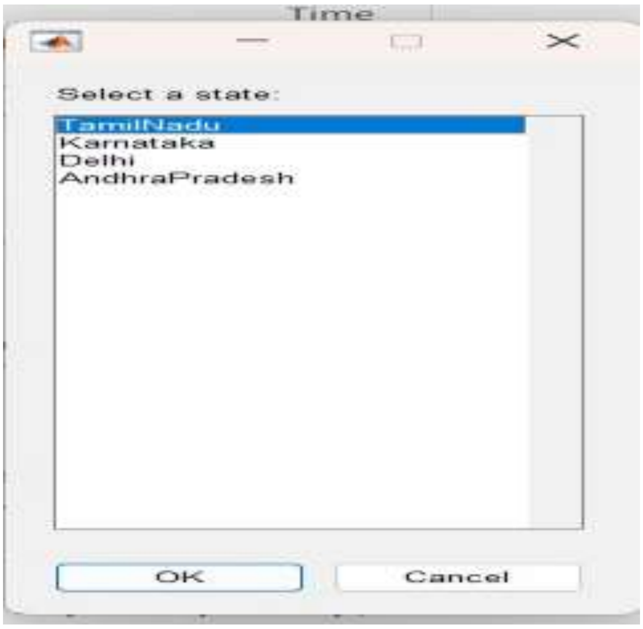
The system utilizes air quality data from Kaggle, encompassing key parameters such as PM2.5, CO, NO2, and SO2. The data undergoes rigorous preprocessing to handle missing values, normalize features, and prepare it for model input. A CNN is employed to extract spatial features from the preprocessed data, identifying intricate relationships between different air quality parameters. Subsequently, an LSTM network analyzes the temporal sequences of these features, capturing the dynamic evolution of air quality over time.

The outputs of the CNN and LSTM components are integrated to create a robust hybrid model capable of delivering accurate air quality forecasts. The entire system is implemented in MATLAB, providing a platform for model development, training, and visualization of prediction results. The model's performance is evaluated using rigorous evaluation metrics, demonstrating its superiority in predicting air quality levels compared to conventional methods.

5. MODULES

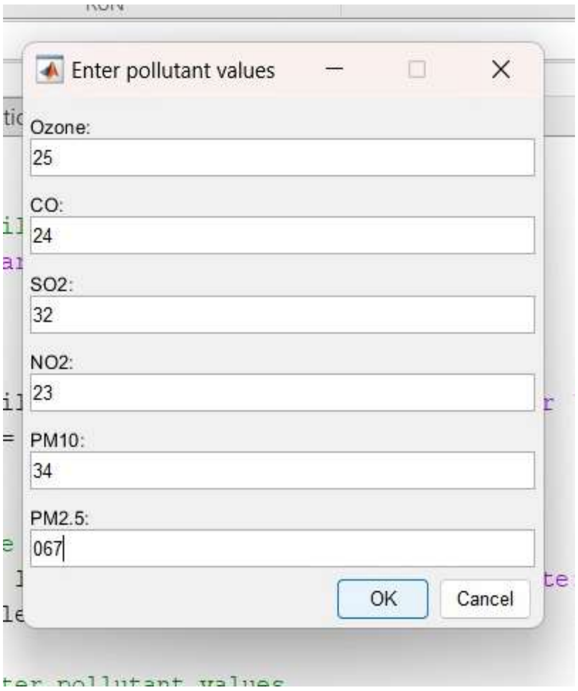
- Data Preprocessing Module: Responsible for cleaning, transforming, and preparing the input data for model training. Includes functions for handling missing values, normalizing sequences /standardizing data, and creating time-series.
- CNN Feature Extraction Module: Implements a Convolutional Neural Network (CNN) to extract spatial features from the preprocessed data. Includes convolutional layers, pooling layers, and activation functions.
- LSTM Temporal Analysis Module: Implements a Long Short-Term Memory (LSTM) network to analyze the temporal sequences of the features extracted by the CNN. Includes LSTM layers, dropout layers, and dense layers.
- Hybrid Model Integration Module: Combines the outputs of the CNN and LSTM modules to create a hybrid model. Includes concatenation or fusion layers to integrate the features.
- Prediction Module: Generates air quality predictions based on the trained hybrid model. Includes functions for making predictions and formatting the output.
- Visualization Module: Creates visualizations of predicted and actual air quality levels, performance metrics, and feature importance. Utilizes MATLAB's visualization tools.
- Evaluation Module: Calculates and displays the models performance metrics.

6. RESULT


Select a state:

TamilNadu
Karnataka
Delhi
AndhraPradesh

Figure 3. Select the State


Enter pollutant values

Ozone: 25

CO: 24

SO2: 32

NO2: 23

PM10: 34

PM2.5: 067

Figure 4. Enter the Values of Pollutant Air

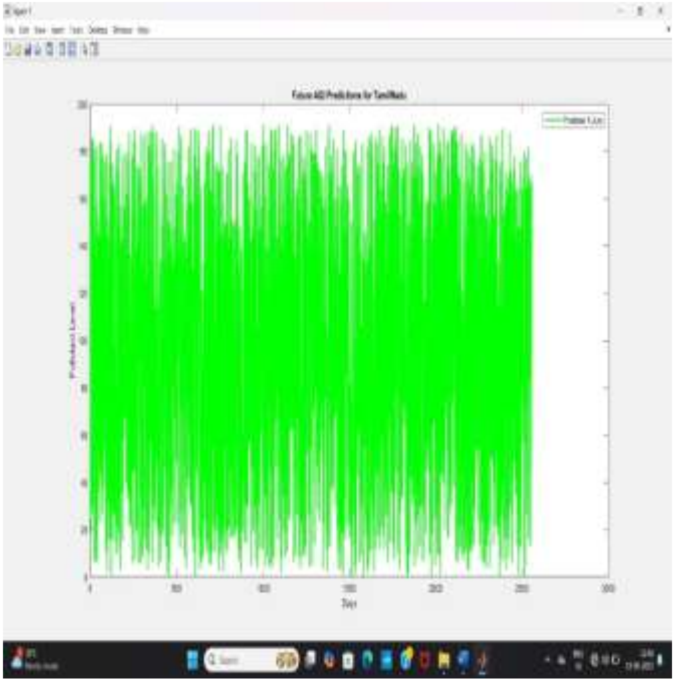


Figure 5. AQI Prediction for Tamilnadu

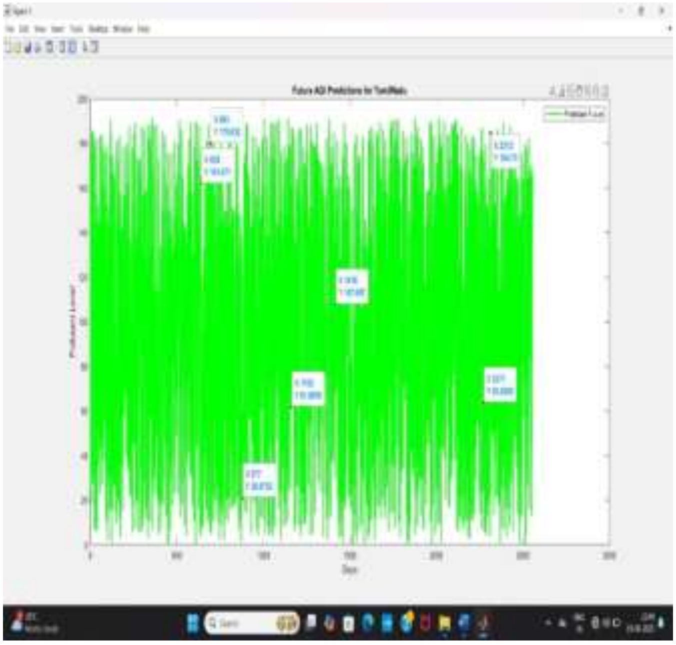


Figure 6. AQI Prediction with Days

7. CONCLUSIONS

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. 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 H2MCM model in predicting air quality is well-established, as evident from its impressive PLCC and RMSE metrics. Through experimentation, we have demonstrated that the H2MCM model surpasses SVR, MLP, RAQP, and Vlachogianni models by 75%, 60%, 38%, and 70%, respectively. However, we acknowledge that while the H2MCM model exhibits exceptional performance, it may not always be the optimal choice for every scenario.

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. 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.
5. A. Vlachogianni, P. Kassomenos, A. Karppinen, S. Karakitsios, and J. Kukkonen, ‘‘Evaluation of a multiple regression model for the forecast- ing 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.
6. *Weather Dataset*. Accessed: Jul. 15, 2023. [Online]. Available: https://www.kaggle.com/datasets/muthuj7/weather-dataset
7. 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.
8. 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.

9. 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.
10. 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
11. 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.
12. 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.
13. 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.
14. 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.
15. N. Geneva and N. Zabaras, ''Transformers for modeling physical systems,''
Neural Netw., vol. 146, pp. 272–289, Feb. 2022