Letters in Information Technology Education (LITE)
Vol 7, No 2, 2024, pp. 63-69
e-ISSN: 2654-5667

Improving Urban Air Quality Prediction Using Bidirectional GRU: A Case Study of CO Concentration to Support Education in Yogyakarta

I Wayan Ordiyasa[a,1], Sriwidodo[b,2], Harits Dwi Wiratmaa[,3], Mohammad Diqi[a,4,*], Marselina Endah Hiswati[a,5], Noverianus[a,6], Namira Anjani Rahmadina Syihab[b,7]
[a] Universitas Respati Yogyakarta, Jl. Laksda Adisucipto KM.6,3 Yogyakarta 55281 Indonesia
[b] Universitas Respati Indonesia, Jl. Bambu Apus I, No. 33, Daerah Khusus Ibukota Jakarta 13890, Indonesia
[1] wayanordi@respati.ac.id; [2] sriwidodo@urindo.ac.id; [3] harits@respati.ac.id; [4] diqi@respati.ac.id*;
[5] marsel.endah@respati.ac.id; [6] 22220027@respati.ac.id; [7] 225050045@urindo.ac.id
* corresponding author

Article Info

Article history:
Received: Nov 5, 2024
Revised: Dec 7, 2024
Accepted: Dec 9, 2024

Keyword:
Air quality prediction
Bidirectional gru
Carbon monoxide
Environmental education

ABSTRACT

Urban air pollution, mainly carbon monoxide (CO), poses significant health risks. This research introduces an educational framework through the application of a Bidirectional Gated Recurrent Unit (Bi-GRU) model for predicting CO levels. The study highlights the model's ability to forecast CO concentrations in Yogyakarta, providing practical insights to support environmental education and awareness among students, researchers, and policymakers. With input lengths of 48, 96, and 144 hours, the model achieved optimal results with an R^2 of 0.99, demonstrating its reliability in capturing CO fluctuations. These findings not only advance machine learning applications for air quality monitoring but also serve as a valuable tool for integrating environmental topics into educational curricula.

I. INTRODUCTION

Air quality monitoring, particularly of carbon monoxide (CO) levels, is crucial in addressing public health risks in urban areas [1]. CO is a colourless, odorless gas released mainly from vehicle emissions and industrial activities, contributing to a significant share of air pollution in cities [2], [3]. According to the World Health Organization, high CO exposure can lead to severe cardiovascular and respiratory problems, affecting vulnerable populations such as children and the elderly [4]. In densely populated areas, CO levels can surpass safe limits, causing thousands of premature deaths annually [5]. Thus, continuous monitoring of CO levels is essential to mitigate health risks and guide environmental policies for cleaner air [6].

The impact of CO pollution on public health has been well-documented, showing a clear correlation between air pollution spikes and hospital admissions [7]. For example, in major cities worldwide, increases in CO levels are linked with heightened emergency visits for respiratory distress [8]. With urbanisation and industrialization on the rise, effective CO monitoring is imperative to protect populations from escalating health hazards [9]. Cities implementing advanced air quality monitoring can promptly detect and address hazardous CO levels, potentially reducing the burden of disease and improving residents' quality of life [10]. Therefore, the prioritisation of CO monitoring in public health frameworks is a proactive step toward sustainable urban living and population well-being [11].

Machine learning models have become instrumental in enhancing the prediction accuracy of pollutant levels within environmental applications. Research shows that ensemble learning techniques can significantly improve predictions in ecological science, though these models demand substantial computational resources [13]. Multi-task learning and regularisation techniques further refine predictive capabilities, especially in hourly air pollution predictions, despite being limited by training data quality and specific regularisation choices [14]. Additionally, the ANFIS-WELM model, which integrates an Adaptive Neuro-Fuzzy Inference System with a Weighted Extreme Learning Machine, provides high accuracy for multi-step pollutant predictions and excels in real-time performance [15]. However, this model's effectiveness is hindered by the complexity of managing input fuzziness, indicating a trade-off between accuracy and complexity.

Integrating machine learning with Chemical Transport Models (CTM) has been shown to improve gridded prediction accuracy, mainly through techniques like the Ensemble Kalman Filter and CTM Fusion. This approach leverages machine learning and CTM to enable 3D spatial predictions, though it is constrained by a dependency on ensemble CTM for error covariance adjustments [16]. A hybrid ensemble model using stacking techniques further enhances the accuracy of air pollution forecasts over time by combining multiple predictive models despite its increased complexity and high data processing demands [17]. Additionally, the CNN+LSTM hybrid deep learning model excels in predicting pollutant concentrations across multiple locations with spatial-temporal precision, though it requires fine-tuning parameters specific to each location for optimal accuracy [18]. Together, these approaches highlight the advancements and challenges in applying machine learning to environmental predictions in complex, dynamic systems.

Feature selection has proven effective in enhancing accuracy for multi-pollutant predictions, utilising techniques like Multi-Target Regression and Feature Ranking to achieve efficient selection with high precision. However, this approach faces challenges due to sensitivity to extreme values and prolonged runtime when applied to large datasets [19]. Additionally, optimised deep learning models, particularly the 1D CNN + LSTM hybrid, demonstrate significant latency reduction, making them well-suited for deployment on resource-limited edge devices. While this model offers promising performance in low-latency applications, its use is currently restricted to PM2.5 predictions and depends on specific hardware capabilities of edge devices [20]. These findings underscore the balance between accuracy, computational demands, and hardware limitations in environmental pollutant prediction.

The gradient-boosted tree model has shown effectiveness in enhancing spatial and temporal modelling for pollutants like NO_2 and O_3, combining accuracy with computational efficiency on large datasets. This model's limitations arise in non-pollutant dense regions, where its effectiveness decreases [21]. Meanwhile, the Kalman-attention-LSTM model has improved predictions for the Air Quality Index (AQI) by incorporating a Kalman filter, attention mechanism, and LSTM, providing higher accuracy and adaptability for AQI and pollutant forecasting. However, this model requires recalibration when applied to different pollutant types, which can limit its versatility [12]. Both approaches demonstrate advancements in predictive accuracy but also highlight the ongoing trade-offs in model adaptability and specificity for diverse environmental conditions.

Previous models for air quality sensor data often struggle with effectively capturing temporal patterns due to limitations in handling complex time series data. Standard machine learning models, such as linear regression or basic neural networks, lack the sophistication to accurately model long-term dependencies in pollutant levels, leading to suboptimal predictions, especially in fluctuating environments [13][14]. Additionally, many existing approaches require extensive computational resources, which limits their practicality for real-time air quality monitoring [15]. The Bidirectional Gated Recurrent Unit (Bi-GRU) offers a promising alternative by processing time series data in both forward and backward directions, allowing for a more comprehensive understanding of temporal dependencies [16]. This bidirectional approach enables Bi-GRU to achieve higher prediction accuracy, making it better suited for capturing intricate patterns in air quality sensor data, ultimately improving real-time pollution management [17].

This research aims to improve the accuracy of carbon monoxide (CO) concentration predictions by leveraging an optimised Bidirectional Gated Recurrent Unit (Bi-GRU) model designed specifically for time-series air quality data. By processing CO levels in both forward and backward directions, the Bi-GRU captures intricate temporal patterns, surpassing traditional models in its ability to predict short- and medium-term fluctuations. Beyond advancing predictive accuracy, this study contributes to education by offering a practical, data-driven approach that can be integrated into environmental education curricula. By engaging students, young researchers, and policymakers, the research emphasises the importance of air quality monitoring and its direct impact on public health and urban sustainability. Through hands-on learning with the Bi-GRU model, learners can bridge theory and practice, fostering a deeper understanding of how advanced technology can address real-world challenges in environmental science.

II. METHOD

A. Dataset

The dataset utilised in this study was obtained from Kaggle and comprises hourly carbon monoxide (CO) measurements gathered by chemical air quality sensors positioned in Yogyakarta, Indonesia. Covering an entire year, from January 1, 2021, to December 31, 2021, the data offers a detailed perspective on CO fluctuations throughout different seasons and times of day, as illustrated in Fig. 1. The sensors are strategically distributed across the city to account for spatial variations in CO levels, capturing differences influenced by location and traffic density. For analytical consistency, each CO measurement was transformed into the Pollutant Standards Index (PSI), simplifying the interpretation of pollutant severity. Data preprocessing steps, including cleaning and min-max normalisation, were applied to enhance the quality and consistency of the dataset, ensuring robust and reliable inputs for the predictive model.

B. Bi-GRU architecture

The Bi-GRU architecture in this study is designed to handle time-series data with various sequence lengths, allowing it to capture both short- and medium-term temporal patterns in CO levels. The model takes inputs of 48, 96, and 144-hour sequences and predicts outputs for the next 24 and 48 hours. This structure enables the model to learn dependencies over different periods, adapting to both immediate and gradual changes in CO concentrations. The Bi-GRU consists of two layers, each with a hidden state dimension of 128 units, providing sufficient capacity for modelling complex temporal dependencies while maintaining computational efficiency.

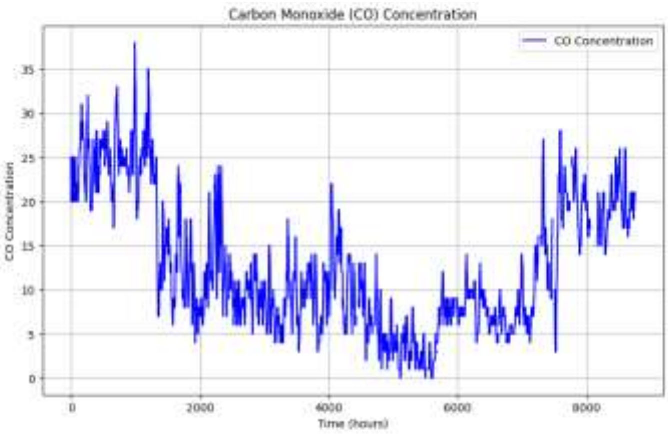

Fig. 1. Carbon Monoxide (CO) Dataset

In a GRU cell, the calculations for the hidden state h_t at time step t are as follows. Update Gate z_t and Reset Gate r_t are formulated in Equations (1) and (2).

$$z_t = \sigma(W_z \cdot x_t + U_z \cdot h_{t-1} + b_z) \tag{1}$$

$$r_t = \sigma(W_r \cdot x_t + U_r \cdot h_{t-1} + b_r) \tag{2}$$

Candidate Activation $\tilde{h}_t$ is calculated in Equation (3).

$$\tilde{h}_t = \tanh(W_h \cdot x_t + U_h \cdot (r_t \odot h_{t-1}) + b_h) \tag{3}$$

Final Hidden State h_t is formulated in Equation (4).

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \tag{4}$$

For a bidirectional GRU (Bi-GRU), the hidden states from the forward ($h_t^{\rightarrow}$) and backward ($h_t^{\leftarrow}$) passes are concatenated at each time step as calculated in Equation (5).

$$h_t^{\mathrm{B\iota-\Gamma PY}} = [h_t^{\rightarrow}; h_t^{\leftarrow}] \tag{5}$$

where [;] denotes concatenation. This bidirectional setup allows the model to leverage information from both past and future contexts, enhancing its ability to capture complex temporal patterns in air quality data.

C. Evaluation Metrics

The evaluation metrics used in this study provide a comprehensive assessment of the model's prediction accuracy for carbon monoxide (CO) concentration levels.

Root Mean Squared Error (RMSE). RMSE measures the square root of the average squared differences between predicted and actual values. It emphasizes larger errors due to squaring, making it sensitive to outliers. RMSE is calculated in Equation 6.

$$\mathrm{PM\Sigma E} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{6}$$

where $\hat{y}_i$ is the predicted value, y_i is the actual value and n is the total number of observations.

Mean Absolute Error (MAE). MAE calculates the average of the absolute differences between predicted and actual values, providing an interpretable measure of error in the same unit as the target variable. It is less sensitive to outliers than RMSE and is calculated in Equation (7).

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \tag{7}$$

Mean Absolute Percentage Error (MAPE). MAPE expresses the prediction error as a percentage, making it easier to interpret in terms of relative error. However, MAPE can be skewed if actual values are close to zero. It is calculated in Equation (8).

$$\mathrm{MA\Pi E} = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right| \tag{8}$$

Coefficient of Determination (R^2). R^2, or the goodness-of-fit measure, represents the proportion of the variance in the actual values that is predictable from the model. An R^2 value closer to 1 indicates a better fit. It is calculated in Equation (9).

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{9}$$

where $\bar{y}$ is the mean of the actual values.

These metrics together provide a well-rounded view of the model's predictive performance, covering both error magnitude (RMSE and MAE), relative accuracy (MAPE), and goodness of fit (R^2).

D. Training and Validation

The dataset in this study is divided into three parts: 70% for training, 20% for validation, and 10% for testing. This split allows the model to learn effectively from a substantial portion of the data, while using the validation set to fine-tune hyperparameters and avoid overfitting. For training, a learning rate of 0.001 is used, combined with the Adam optimiser, which is well-suited for handling sparse gradients and adapting the learning rate dynamically. The batch size is set to 64, balancing computational efficiency and model stability by updating the model's weights after every 64 samples. This approach ensures a robust and well-generalized model, tested on unseen data, with the goal of achieving high accuracy in predicting CO concentration levels in air quality data.

III. Result and Discussion

A. Performance Metrics

Table 1 presents the model's performance across different input-output sequence lengths, highlighting its accuracy in predicting CO levels. As the input length increases, there is a notable improvement in error metrics such as MSE, RMSE, and MAE, with the best results achieved at 144-hour inputs, reflecting the model's ability to capture more complex temporal patterns. For instance, with a 144-hour input and a 24-hour output, the model achieves an MSE of 0.32, an RMSE of 0.57, and an MAE of 0.38, demonstrating excellent precision with an R^2 of 0.99. Similarly, the lowest MAPE values are observed with longer input sequences, indicating minimal deviation from actual values and strong reliability for practical applications. These findings underscore the importance of longer input sequences in enhancing predictive accuracy, especially for real-time air quality monitoring in urban settings.

TABLE I. Performance Metrics

Input	Output	MSE	RMSE	MAE	MAPE	R^2
48	24	0.40	0.63	0.42	0.05%	0.98
96	24	0.35	0.59	0.40	0.04%	0.99
144	24	0.32	0.57	0.38	0.03%	0.99
48	48	0.45	0.67	0.44	0.06%	0.97
96	48	0.38	0.62	0.41	0.04%	0.98
144	48	0.33	0.57	0.39	0.03%	0.99

Table 1 illustrates that as input sequence length increases, the model consistently delivers higher accuracy in predicting CO levels, particularly with a 144-hour input window. This trend suggests that longer input sequences enable the model to capture more temporal dependencies, leading to more precise forecasts. The low MSE, RMSE, and MAE values, coupled with an R^2 of 0.99 for both 24-hour and 48-hour outputs, indicate that the model effectively generalises and adapts to varying prediction horizons. However, the slightly lower R^2 and higher error metrics for the 48-hour output with shorter input sequences imply that capturing longer dependencies is crucial for maintaining performance, significantly when extending prediction lengths. This pattern highlights the need for substantial historical data input to achieve optimal performance, aligning with the model's bidirectional structure, which thrives on comprehensive temporal information.

Despite these strengths, the model's reliance on longer input sequences may pose challenges in practical deployment, especially in settings with limited data storage or processing capabilities. Additionally, the performance difference between input lengths suggests that the model might not fully utilise shorter sequences, which could restrict its adaptability in real-time monitoring scenarios with rapid data updates. Another consideration is that while longer inputs yield higher accuracy, they may also introduce potential noise, impacting stability in environments with high pollutant fluctuations. Future work could explore hybrid models that balance input length with filtering mechanisms to mitigate noise effects while maintaining accuracy. Overall, Table 1 emphasises the importance of tailored input-output configurations to maximise the Bi-GRU model's predictive capabilities for air quality applications.

B. Training and Validation Loss

Fig. 2 illustrates the training and validation loss trajectories over 100 epochs, showcasing the model's learning progression and stability. This result is based on Input-Output 144-24. Initially, both losses start relatively high, with a rapid decrease observed within the first few epochs, indicating effective initial learning. As training continues, the losses stabilise at a low level, demonstrating that the model has achieved a firm fit on the training data without significant overfitting. The slight fluctuations in validation loss reflect minor adjustments as the model adapts to unseen data, maintaining generalisation ability. Overall, the convergence of training and validation losses suggests that the model has been well-tuned and is effective in predicting CO levels with minimal error.

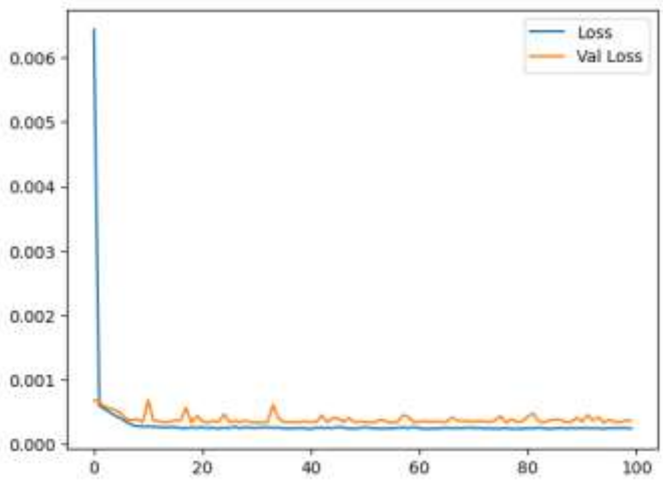

Fig. 2. Training and validation loss

The results shown in Fig. 2 suggest that the model has effectively learned to predict CO levels, as indicated by the rapid drop in both training and validation loss during the initial epochs. This sharp decline reflects the model's ability to capture core patterns in the data early on, a promising sign of efficient training. However, the convergence of training and validation losses at a relatively low level also implies that further learning gains are minimal as the model reaches a plateau. This stabilisation could indicate that the model has captured most of the underlying structure in the data, with little room for improvement. The lack of divergence between training and validation loss suggests that overfitting is not a significant issue, indicating a well-generalized model.

Despite these positive outcomes, a few critical observations warrant attention. The minor fluctuations in validation loss suggest sensitivity to specific data patterns that the model encounters during validation, pointing to potential areas of complexity that the model might not fully capture. This residual variability may reflect noise in the data or limitations in the model's capacity to generalise to all temporal variations in CO levels. Additionally, while convergence is generally favourable, it suggests that the model has reached its performance ceiling under the current configuration, leaving little room for improvement without additional tuning or model adjustments. Overall, while the model performs robustly, these findings underscore the importance of continual evaluation and potential refinement to enhance predictive accuracy further.

C. *Comparison of Original and Predicted Data*

Fig. 3 compares the original CO concentration data with the model's predicted values, illustrating the model's effectiveness in tracking the actual data trends. This result is based on Input-Output 144-24. The close alignment between the two lines suggests that the model accurately captures both the overall pattern and the fluctuations in CO levels over time. Minor deviations are visible but appear minimal, indicating a high level of precision in the model's predictions. This strong correlation implies that the model has effectively learned the underlying structure of the time-series data, allowing it to forecast CO concentrations reliably. Overall, Fig. 3 demonstrates the model's potential as a valuable tool for real-time air quality monitoring, closely mirroring actual pollutant behaviour.

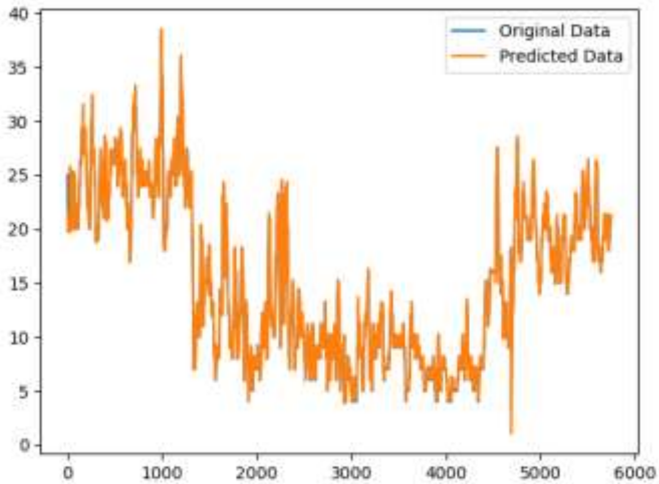

Fig. 3. Comparison of Original and Predicted Data

Fig. 3 reveals that the model performs well in predicting CO levels, with predicted values closely following the original data's trends. This close alignment suggests that the model effectively captures both seasonal variations and short-term fluctuations in CO concentrations, which is crucial for real-time monitoring applications. The occasional minor discrepancies between the original and predicted values highlight areas where the model might slightly under- or overestimate concentrations, possibly due to noise in the data or abrupt environmental changes that the model cannot fully capture. Nonetheless, these deviations are relatively small, demonstrating the model's robustness and its ability to generalise well across the dataset. Overall, the figure underscores the model's reliability and responsiveness to complex temporal patterns in air quality data.

Despite its strengths, a critical examination of Fig. 3 raises questions about the model's performance during peak CO concentrations. In some of these peaks, the model appears to smooth out extreme values, indicating a potential limitation in accurately capturing sudden spikes in pollutant levels. This smoothing effect, while not significantly detrimental, could limit the model's effectiveness in scenarios where extreme values are vital for health risk assessments and immediate responses. Additionally, the strong alignment observed here might be dataset-specific, suggesting the need for further testing on diverse datasets to confirm the model's generalizability. Overall, while Figure 3 demonstrates promising results, these observations highlight the importance of continuous model evaluation and adaptation to ensure accurate predictions across varied conditions.

D. *Summarization of Key Findings*

This research addresses the critical challenge of accurately predicting carbon monoxide (CO) levels in urban environments, focusing on enhancing real-time air quality monitoring. Through an optimised Bidirectional Gated Recurrent Unit (Bi-GRU) model, the study demonstrates that longer input sequences significantly improve prediction accuracy, capturing complex temporal dependencies in CO fluctuations. Key findings include a high R^2 of 0.99 for 144-hour inputs, with low MSE, RMSE, and MAE values, indicating the model's substantial fit and minimal deviation from actual values. The results reveal that extending input length enhances the model's precision, particularly for both 24-hour and 48-hour prediction windows. These findings underscore the Bi-GRU model's potential as a reliable tool for urban air quality forecasting, with implications for improved public health and environmental management.

E. *Result Interpretations*

The results show a clear pattern where increased input sequence length leads to improved prediction accuracy, with the model performing best at 144-hour inputs for both 24-hour and 48-hour prediction windows. This relationship suggests that the Bi-GRU model effectively leverages extended temporal information, aligning well with the initial expectation that longer input sequences capture complex trends in CO fluctuations. However, slight performance drops in shorter input sequences, especially for 48-hour predictions, were unexpected and may indicate limitations in the model's ability to generalise with minimal historical data. An alternative explanation for this discrepancy could be that shorter sequences fail to provide sufficient context, leading the model to over-rely on immediate past values rather than capturing broader trends. Overall, these insights affirm the value of longer sequences but highlight the need for further exploration of model adjustments to improve performance with shorter inputs.

F. *Research Implications*

This research holds significant potential not only for advancing air quality prediction in urban environments but also for enriching educational practices in environmental science. By demonstrating the effectiveness of the Bi-GRU model in capturing temporal dependencies in CO levels, the study aligns with prior findings on the value of advanced neural networks over traditional methods for handling complex time-series data. The insights gained, such as the impact of longer input sequences on model performance, provide not just technical advancements but also a valuable educational tool. The Bi-GRU model can serve as a learning framework for students to explore data analysis, machine learning, and the real-world implications of air pollution. By incorporating the case study of Yogyakarta into academic settings, students can bridge theory with practice, gaining a deeper understanding of the critical role of technology in environmental monitoring and public health. This integration underscores the dual impact of the research: fostering innovation in predictive modelling while empowering future generations to address pressing environmental challenges.

G. Research Limitations

This study concludes that the Bi-GRU model is highly effective for predicting CO levels in urban air quality monitoring, particularly with longer input sequences. However, a limitation is the model's decreased accuracy with shorter input sequences, which may hinder its adaptability in real-time settings with limited historical data. Despite this, the results remain robust, as the model consistently performed well across the majority of scenarios, achieving high R^2 values and low error metrics. These limitations highlight potential areas for refinement but do not detract from the model's core capability of accurately capturing temporal patterns in CO data. Thus, the findings are still valid and effectively address the research question, affirming the Bi-GRU model's utility in enhancing air quality prediction systems.

H. Recommendations for Future Research

For practical implementation, it is recommended to incorporate the Bi-GRU model into urban air quality monitoring systems, particularly in areas with high pollution variability, to improve real-time CO prediction and support rapid public health responses. Future research should investigate ways to enhance the model's performance with shorter input sequences, which would improve adaptability in real-time applications where extensive historical data may not be available. Additionally, exploring hybrid models that combine Bi-GRU with attention mechanisms or data filtering techniques could help capture critical pollution spikes without compromising accuracy. Researchers could also evaluate the model's performance across different pollutants and geographic locations to test its generalizability and robustness. These efforts would refine the Bi-GRU's effectiveness and broaden its applicability in comprehensive environmental monitoring frameworks.

IV. Conclusion

This study successfully demonstrates the effectiveness of the Bidirectional Gated Recurrent Unit (Bi-GRU) model in predicting urban carbon monoxide (CO) levels, with a focus on Yogyakarta as a case study. By leveraging its bidirectional architecture, the model captures intricate temporal patterns in air quality data, achieving remarkable accuracy with a coefficient of determination (R^2) of 0.99 for longer input sequences. These results highlight the model's potential as a reliable tool for real-time air quality monitoring, addressing critical urban challenges posed by CO pollution. Beyond its technical contributions, this research offers practical insights for integrating advanced machine learning applications into environmental education. By empowering students, young researchers, and policymakers with data-driven tools, this work underscores the importance of technology in fostering environmental awareness and sustainable urban living. While the Bi-GRU model excels in capturing complex dependencies in extensive datasets, its reduced performance with shorter input sequences presents an opportunity for future improvements. This limitation invites further exploration into hybrid approaches that combine Bi-GRU with filtering or attention mechanisms to enhance adaptability in real-time scenarios. Additionally, expanding the model's application to other pollutants and geographic contexts could validate its generalizability and broaden its impact. These advancements would refine the model's utility as a critical resource for addressing pollution-related health risks and improving urban air quality. Ultimately, this research bridges technological innovation and human-centred goals, aligning predictive precision with the urgent need for healthier, more sustainable communities.

References

[1] R. K. Angatha and A. Mehar, "Impact of Traffic on Carbon Monoxide Concentrations Near Urban Road Mid-Blocks," *J. Inst. Eng. India Ser. A*, vol. 101, no. 4, pp. 713–722, 2020, doi: 10.1007/s40030-020-00464-2.

[2] H. E. Obanya, N. H. Amaeze, O. Togunde, and A. A. Otitoloju, "Air Pollution Monitoring Around Residential and Transportation Sector Locations in Lagos Mainland," *J. Health Pollut.*, vol. 8, no. 19, 2018, doi: 10.5696/2156-9614-8.19.180903.

[3] I. Gregorczyk-Maga *et al.*, "Air Pollution May Affect the Assessment of Smoking Habits by Exhaled Carbon Monoxide Measurements," *Environ. Res.*, vol. 172, pp. 258–265, 2019, doi: 10.1016/j.envres.2019.01.063.

[4] Y. Wang *et al.*, "The Impact of Carbon Monoxide on Years of Life Lost and Modified Effect by Individual- And City-Level Characteristics: Evidence From a Nationwide Time-Series Study in China," *Ecotoxicol. Environ. Saf.*, vol. 210, p. 111884, 2021, doi: 10.1016/j.ecoenv.2020.111884.

[5] K. Chen *et al.*, "Ambient Carbon Monoxide and Daily Mortality: A Global Time-Series Study in 337 Cities," *Lancet Planet. Health*, vol. 5, no. 4, pp. e191–e199, 2021, doi: 10.1016/s2542-5196(21)00026-7.

[6] A. Rebeiro-Hargrave *et al.*, "City Wide Participatory Sensing of Air Quality," *Front. Environ. Sci.*, vol. 9, 2021, doi: 10.3389/fenvs.2021.773778.

[7] N. A. Manan, A. N. Aizuddin, and R. Hod, "Effect of Air Pollution and Hospital Admission: A Systematic Review," *Ann. Glob. Health*, vol. 84, no. 4, p. 670, 2018, doi: 10.29024/aogh.2376.

[8] A. Phosri, K. Ueda, V. L. H. Phung, B. Tawatsupa, A. Honda, and H. Takano, "Effects of Ambient Air Pollution on Daily Hospital Admissions for Respiratory and Cardiovascular Diseases in Bangkok, Thailand," *Sci. Total Environ.*, vol. 651, pp. 1144–1153, 2019, doi: 10.1016/j.scitotenv.2018.09.183.

[9] X. Li *et al.*, "The Impact of Ambient Air Pollution on Hospital Admissions, Length of Stay and Hospital Costs for Patients With Diabetes Mellitus and Comorbid Respiratory Diseases in Panzhihua, Southwest China," *J. Glob. Health*, vol. 13, 2023, doi: 10.7189/jogh.13.04118.

[10] A. F. W. Ho *et al.*, "Ambient Air Quality and Emergency Hospital Admissions in Singapore: A Time-Series Analysis," *Int. J. Environ. Res. Public. Health*, vol. 19, no. 20, p. 13336, 2022, doi: 10.3390/ijerph192013336.

[11] Y. Tian *et al.*, “Ambient Air Pollution and Daily Hospital Admissions: A Nationwide Study in 218 Chinese Cities,” *Environ. Pollut.*, vol. 242, pp. 1042–1049, 2018, doi: 10.1016/j.envpol.2018.07.116.

[12] H. Zhou, T. Wang, H. Zhao, and Z. Wang, “Updated Prediction of Air Quality Based on Kalman-Attention-LSTM Network,” *Sustainability*, vol. 15, no. 1, p. 356, 2022, doi: 10.3390/su15010356.

[13] A. Masih, “Machine learning algorithms in air quality modelling,” *Glob. J. Environ. Sci. Manag.*, vol. 5, no. 4, pp. 515–534, 2019, doi: 10.22034/GJESM.2019.04.10.

[14] D. Zhu, C. Cai, T. Yang, and X. Zhou, “A Machine Learning Approach for Air Quality Prediction: Model Regularization and Optimization,” *Big Data Cogn. Comput.*, vol. 2, no. 1, p. 5, 2018, doi: 10.3390/bdcc2010005.

[15] Y. Li, P. Jiang, Q. She, and G. Lin, “Research on Air Pollutant Concentration Prediction Method Based on Self-Adaptive Neuro-Fuzzy Weighted Extreme Learning Machine,” *Environ. Pollut.*, vol. 241, pp. 1115–1127, 2018, doi: 10.1016/j.envpol.2018.05.072.

[16] L. Fang *et al.*, “A Gridded Air Quality Forecast Through Fusing Site-Available Machine Learning Predictions From RFSML v1.0 and Chemical Transport Model Results From GEOS-Chem v13.1.0 Using the Ensemble Kalman Filter,” *Geosci. Model Dev.*, vol. 16, no. 16, pp. 4867–4882, 2023, doi: 10.5194/gmd-16-4867-2023.

[17] Y.-S. Chang, S. Abimannan, H.-T. Chiao, C.-Y. Lin, and Y.-P. Huang, “An Ensemble Learning Based Hybrid Model and Framework for Air Pollution Forecasting,” *Environ. Sci. Pollut. Res.*, vol. 27, no. 30, pp. 38155–38168, 2020, doi: 10.1007/s11356-020-09855-1.

[18] A. Gilik, A. S. Öğrenci, and A. Özmen, “Air Quality Prediction Using CNN+LSTM-based Hybrid Deep Learning Architecture,” *Environ. Sci. Pollut. Res.*, vol. 29, no. 8, pp. 11920–11938, 2021, doi: 10.1007/s11356-021-16227-w.

[19] S. Masmoudi, H. Elghazel, D. Taieb, O. Yazar, and A. Kallel, “A Machine-Learning Framework for Predicting Multiple Air Pollutants’ Concentrations via Multi-Target Regression and Feature Selection,” *Sci. Total Environ.*, vol. 715, p. 136991, 2020, doi: 10.1016/j.scitotenv.2020.136991.

[20] I. N. K. Wardana, J. W. Gardner, and S. A. Fahmy, “Optimising Deep Learning at the Edge for Accurate Hourly Air Quality Prediction,” *Sensors*, vol. 21, no. 4, p. 1064, 2021, doi: 10.3390/s21041064.

[21] V. Balamurugan, J. Chen, A. Wenzel, and F. N. Keutsch, “Spatiotemporal Modeling of Air Pollutant Concentrations in Germany Using Machine Learning,” *Atmospheric Chem. Phys.*, vol. 23, no. 17, pp. 10267–10285, 2023, doi: 10.5194/acp-23-10267-2023.