International Journal of Combinatorial Optimization Problems and Informatics, 17(3), 2026, 71-81. ISSN: 2007-1558.
https://doi.org/10.61467/2007.1558.2026.v17i3.897

www.editada.org

Enhanced CO_2 Forecasting in Indoor Environments Using Advanced LSTM Models

Ariel Isaac Posada Barrera [1], *Laura Margarita Rodríguez Peralta* [1,2]

[1] Department of Engineering, Faculty of Information Technology and Data Science (FTIyCD), Universidad Popular Autónoma del Estado de Puebla (UPAEP), Puebla, México

[2] Secretaría de Ciencia, Humanidades, Tecnología e Innovación (SECIHTI), Puebla, México

arielisaac.posada@upaep.edu.mx, lauramargarita.rodriguez01@upaep.mx

Abstract. Accurate forecasting of CO_2 levels in indoor environments is essential for effective air quality management and public health protection. This study evaluates the performance of four Long Short-Term Memory (LSTM)-based models—LSTM, Scalar LSTM (sLSTM), Matrix LSTM (mLSTM), and Extended LSTM (xLSTM)—for CO_2 prediction in indoor settings. Results show that xLSTM consistently outperforms the other models across multiple evaluation metrics, establishing it as a robust and reliable approach for air quality monitoring. The research highlights the importance of precise CO_2 forecasting for policymakers and building managers in optimizing indoor environmental conditions. Future work will focus on integrating xLSTM into a real-time monitoring system powered by Big Data and Internet of Things (IoT) technologies. This system will incorporate multiple forecasting algorithms to enhance predictive accuracy and support respiratory disease prevention through continuous, adaptive air quality management.

Keywords: Long Short-Term Memory, Time Series Prediction, Machine learning, Sick buildings syndrome

Article Info
Received February 20, 2025
Accepted Mar 11, 2026

1 Introduction

Indoor air quality is a key for assessing occupant health by preventing cardiac, respiratory, and neurological conditions, as well as reducing the risk of contracting respiratory diseases. Therefore, ensuring optimal indoor environments is essential not only for individual well-being but also for public health.

In recent years, growing evidence has underscored the importance of proactive management strategies to mitigate indoor air pollutants, particularly carbon dioxide (CO_2), given its direct correlation with human occupancy and activity levels (Gabriel et al., 2024). Achieving optimal indoor air quality requires a comprehensive understanding of pollutants, their chemical behaviors, diffusion patterns, and transportation mechanisms within enclosed spaces characterized by routine human activity. Additionally, empirical knowledge complemented by Machine Learning models is feasible due to the inherent complexity of pollutant dynamics and significant variations in pollutant behaviors across different spaces. Not controlled external environmental conditions, weather variations, and distinct patterns of human activity can influence indoor air quality. Machine learning models provide a solution for accurately forecasting indoor pollutant levels under these complex and varying conditions.

This study, conducted in the context of Mexico, addresses the critical challenge of predicting indoor air quality – specifically carbon dioxide (CO_2) concentrations – through the implementation of the Extended Long Short-Term Memory (Beck et al., 2024) method.

Previous research (Domínguez Portillo et al., 2023) has highlighted the urgent need for continuous monitoring and effective management of indoor pollutants to safeguard public health. In this context, related studies have shown that these variables are effective in defining and characterizing health risk in indoor spaces through fuzzy logic–based approaches (Nunes et al., 2025), and that their assessment can be further enhanced through the use of modern quantum algorithms, as explored in (Sotelo Gómez et al. 2025)

Other studies (Posada Barrera et al., 2023, 2024a, 2024b) have focused on forecasting concentrations of carbon dioxide (CO_2), particulate matter (PM2.5, PM10), and volatile organic compounds (TVOC) by deploying regression-based machine learning algorithms. These models incorporate temperature, relative humidity, occupant density, and ventilation characteristics as input variables, yielding favorable results with the Gradient-Boosted Tree Model (Friedman, 2001).

Additionally, a recent study (Linares Alzamora et al., 2025) explored Kolmogorov-Arnold Networks (Liu et al., 2024) as an alternative modeling approach, demonstrating improved forecasting performance particularly for TVOCs.

The challenge of CO_2 forecasting has been explored in depth, proving that it requires modern algorithms with greater complexity such as Transformers, Informers, Recurrent Neural Networks, and Gated Recurrent Units; with proposed solutions using State-Space Gated Recurrent Unit and Decomposition State-Space Gated Recurrent Unit (Mohammadshirazi et al., 2023), as well as series analysis using Autoregressive Integrated Moving Average (Romero-López et al., 2025).

Across all these studies, CO_2 consistently emerged as the most challenging variable to predict accurately, reflecting the influence of multiple factors such as occupant behavior, ventilation patterns, and fluctuating environmental conditions. Furthermore, research has shown that CO_2 levels vary significantly across locations and seasons due to external, uncontrolled factors (Posada Barrera et al., 2024a). As a result, exploring alternative algorithms became essential, given the critical role of accurate CO_2 forecasts in safeguarding public health.

Previous studies (Dai et al., 2024; Gabriel & Auer, 2023; Park, Seo & Cho, 2023; Wei et al., 2023; Yang, Yuang, Wu, 2022; Zhu et al., 2022) have employed Long Short-Term Memory (Hochreiter & Schmidhuber, 1997) based technologies to predict CO_2 concentrations in indoor spaces—an approach that may demand greater computational resources and a more intricate structure than traditional regression models—yet demonstrates superior efficacy in accommodating seasonal variations, adapting through continuous learning, and ensuring reliability in critical areas.

With its enhanced long-term memory capabilities, Extended Long Short-Term Memory (xLSTM) has proven effective in addressing various time series forecasting challenges, offering continuous improvements in prediction accuracy. Its ability to capture intricate temporal patterns makes it a promising solution for CO_2 forecasting in indoor environments, ultimately contributing to more reliable air quality management and the prevention of respiratory illnesses, including Sick Building Syndrome (SBS).

The contribution of this study is the proposal of a predictive model for indoor air quality assessment in Mexico based on the Extended Long Short-Term Memory (xLSTM) architecture. The study primarily constitutes an empirical benchmarking contribution, as it systematically evaluates the performance of the proposed xLSTM-based model against traditional LSTM architectures and their variants, such as sLSTM and mLSTM, as well as standard machine learning approaches used in previous studies. Secondarily, the study provides methodological validation by demonstrating the suitability and robustness of the xLSTM architecture for modeling complex temporal dynamics associated with indoor pollutant behavior.

This paper is organized into the following sections: Section 2, "Dataset," describes the data used for the study, detailing its collection process, characteristics, and relevance to the experiments. Section 3, "Experimental Procedures," outlines the methodology applied in training and evaluating models based on long short-term memory (LSTM) technologies. Section 4, "Results," presents an in-depth analysis of the model's performance through distinct evaluation metrics, highlighting its effectiveness in handling continuous data. Finally, Section 5, "Conclusions and Directions for Further Research," summarizes the key findings of the study and discusses future research avenues to advance the understanding of air quality dynamics. Specifically, it aims to contribute to the formulation of policies and best practices for maintaining indoor air quality, with particular attention to CO_2 as the most challenging variable to predict.

2 Dataset

This study utilizes a subset of data from a previously conducted experiment (Posada Barrera et al., 2024a), selecting only the last 400 continuous measurements – each representing distinct environmental variables – to evaluate the model's performance. Despite the dataset's increased complexity, the model demonstrated strong predictive capabilities while requiring fewer training samples. The selected data corresponds to a stable phase of the original study, ensuring that the test conditions reflect a realistic yet controlled indoor environment.

The dataset originates from an earlier study in which environmental data was collected inside a 3.2×5×2.8-meter home office located in Morelos, Mexico. Over 43 working hours, the original study recorded 25,888 continuous measurements at 10-second intervals using two primary sensor systems: the Databot (Databot, 2023) and an air quality monitor (Inkbird, 2023).

To ensure measurement accuracy, data from the Databot was cross-validated against readings from the Inkbird sensor, minimizing potential errors and enhancing reliability. No outliers were removed, and all recorded measurements were preserved to maintain data integrity for analysis.

The selected 400 measurements capture a critical transition in indoor air conditions. Initially, CO_2 levels were elevated due to the room being enclosed, with both the door and windows closed. Toward the end of this phase, ventilation improved as the door and windows were opened, allowing for better air circulation. This transition was particularly relevant for assessing the model's ability to adapt to realistic variations in indoor environments.

By focusing on this period, the study ensures that the model is tested under conditions that reflect natural fluctuations in air quality while maintaining a controlled and well-documented setting.

For model development, temperature (°C) and relative humidity (%) were selected as input variables, considering their fluctuations over time and their potential influence on indoor air quality. The target variable for prediction was CO_2 in particles per million (ppm).

The room's temperature ranged from 29.31 °C at its lowest to 32.69 °C at its highest, with an average of 31.09 °C. These readings capture how both outdoor weather and regular indoor activities influenced the ambient conditions. Figure 1 presents a histogram of the temperature measurements.

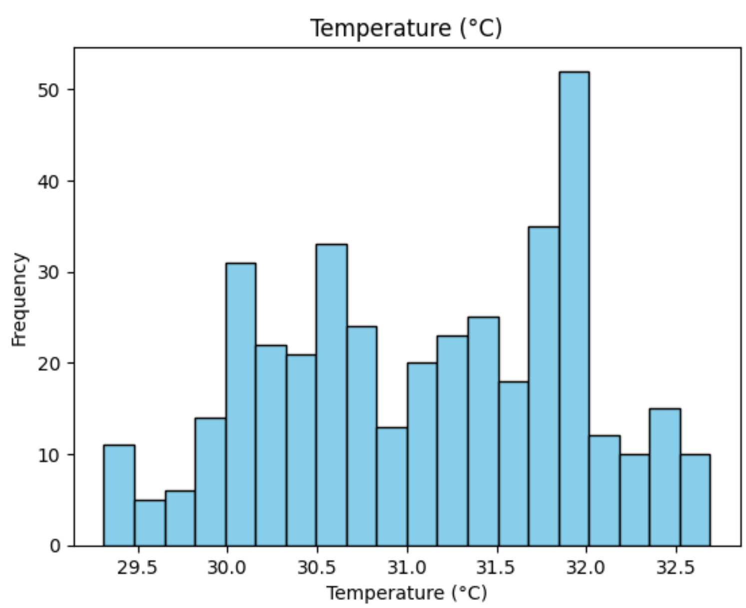

Fig. 3. Temperature histogram.

Meanwhile, relative humidity remained relatively low, varying between 7.98% and 11.17%, with an average of 8.83%. This limited moisture level was likely a result of the regional climate in Morelos, Mexico, combined with the open-air environment in the room. Figure 2 shows a histogram of humidity values.

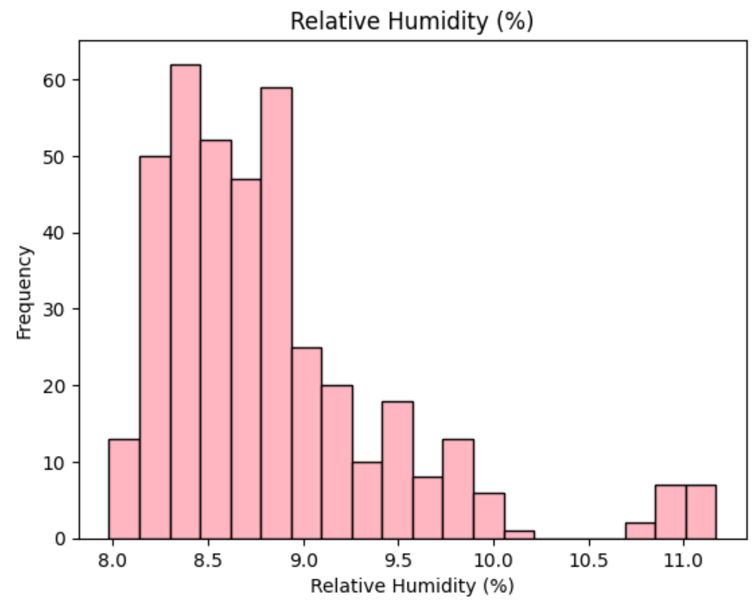

Fig. 2. Relative humidity histogram.

Regarding air quality, CO_2 concentrations ranged from 400 ppm to 693 ppm, with a mean of 458.58 ppm. Although these levels generally fell within a safe range, any prolonged measurements above 600 ppm were noted as potentially contributing to discomfort or heightened respiratory risks. By tracking these fluctuations, the study assessed how everyday conditions impacted indoor air quality. Figure 3 depicts a histogram of the CO_2 measurements.

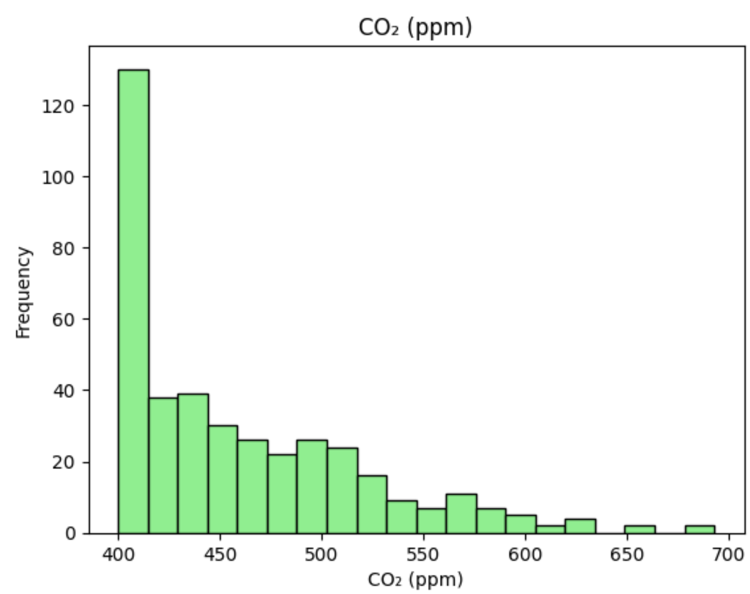

Fig. 3. CO_2 levels histogram.

Using these measurements of relative humidity and temperature, the experiments were conducted with LSTM models to analyze indoor environmental conditions and their impact on air quality.

3 Experimental procedures

To evaluate the capacity of deep learning models to predict CO_2 concentration based on indoor environmental conditions, four architectures based on long short-term memory networks were trained using the selected dataset. The models included:

- Long Short-Term Memory (LSTM)
- Scalar Long Short-Term Memory (sLSTM)
- Matrix Long Short-Term Memory (mLSTM)
- Extended Long Short-Term Memory (xLSTM)

The Long Short-Term Memory (LSTM) network is a type of recurrent neural network (RNN) specifically designed to handle sequences by managing long-range temporal dependencies effectively. LSTM architecture utilize gating mechanisms—input, forget, and output gates—that control the flow of information into and out of a scalar cell state, thus selectively remembering or discarding information over time. Despite their effectiveness, traditional LSTM architectures encounter challenges in scalability and parallelization, especially when handling large datasets or high-dimensional embeddings.

Advanced variants such as scalar Long Short-Term Memory (sLSTM), matrix Long Short-Term Memory (mLSTM), and xLSTM have emerged to overcome these limitations (Beck et al., 2024). The sLSTM retains the scalar memory state characteristic of traditional LSTMs but introduces exponential gating mechanisms, which enable dynamic adjustments in information retention.

Specifically, exponential gates are stabilized using dedicated normalizer states, which accumulate gate activations across future steps to maintain numerical stability and facilitate memory mixing within the scalar structure. Conversely, the mLSTM expands memory capacity by adopting a matrix-based memory state that leverages key-value pair storage. This variant employs a covariance-based update rule, enhancing the storing of information and enabling complete parallelization across sequence positions. This parallelizability improves computational efficiency, making mLSTM highly suitable for large-scale applications.

The xLSTM architecture integrates sLSTM or mLSTM cells into residual blocks, improving residual stacking and high-dimensional embeddings to further enhance scalability and separation of different sequence contexts (Beck et al., 2024).

Table 1. compares the selected models.

Table 1. Comparison of Selected Models

Aspect	xLSTM	LSTM	sLSTM	mLSTM
Architecture & Memory Structure	Stacked residual blocks using sLSTM or mLSTM	Scalar memory cell structure	Scalar memory with normalizer state	Matrix memory with covariance updates
Gating Mechanisms	Exponential gates; enhanced gating via residual stacking	Input, forget, output gates (sigmoid-based)	Exponential input and forget gates	Exponential gates with stabilizing methods
Normalization & Stabilization Techniques	Normalizer state and gate stabilization integrated in blocks	No explicit normalization/stabilization	Normalizer state and exponential gate stabilization	Normalizer state; stabilization of exponential gates to prevent overflow
Parallelizability	Enhanced parallelizability depending on underlying	Limited parallelizability (sequential processing)	Limited due to scalar memory recurrence	Fully parallelizable due to matrix-based memory updates

	architecture			
Memory Mixing Capability	Memory mixing within residual blocks for improved context separation	Basic mixing through recurrent connections	Enhanced mixing within heads using exponential gates	No memory mixing: multiple heads equivalent to multiple cells
Scalability & Performance	Highly scalable with combined architecture	Limited scalability	Improved scalability via stabilized exponential gates and scalar normalization	High scalability through matrix-based associative memory

Multiple experiments were conducted to explore various model configurations, with emphasis on optimizing hyperparameters such as learning rates, epoch counts, and architectural features. The most feasible architecture identified through these experiments utilized a head size of 32 and incorporated two attention heads. These experiments demonstrated improved model performance by effectively capturing temporal dynamics and enhancing the representation of long-term dependencies inherent in the environmental data.

Input features—temperature (°C) and relative humidity (%)—were normalized to standardize values, preventing scale imbalances and ensuring stable convergence during training.

The dataset was divided into 70% for training and 30% for validation, ensuring a representative distribution of environmental conditions across both sets. Training was conducted in an environment equipped with CUDA acceleration, significantly optimizing computational efficiency and reducing processing time for deep learning tasks. Each model was trained for 40 epochs with a learning rate of 0.01, selected after conducting experiments with multiple combinations, balancing sufficient iterations for learning while mitigating the risk of overfitting.

Mean Squared Error (Willmott & Matsuura, 2005) was used as the loss function, ensuring effective optimization by penalizing larger errors and prioritizing more accurate CO_2 predictions. This loss function was chosen for its ability to highlight deviations, reinforcing model adjustments during training. Figure 4 presents the training loss curves for all four models, illustrating their learning progression. Each model exhibited stable convergence, indicating that the training process was successful and that all models achieved an acceptable level of performance.

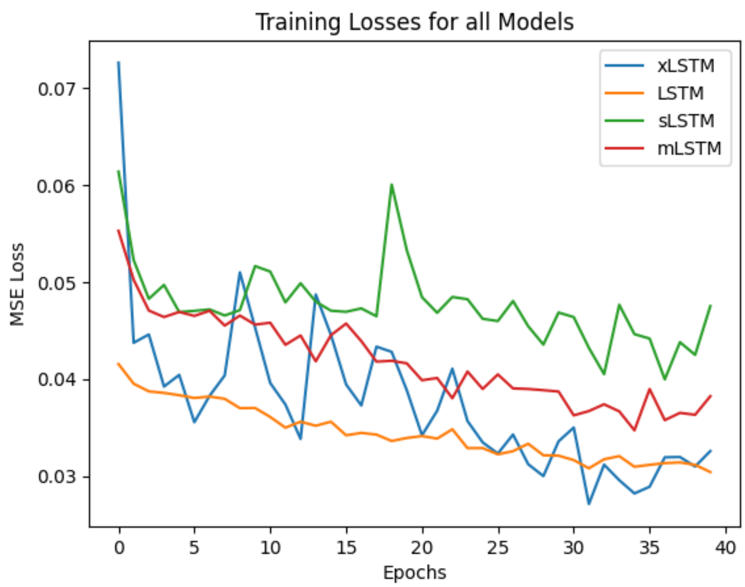

Fig. 4. Training Losses for all Models.

As shown in Figure 4, the two models that demonstrated the best performance were Long Short-Term Memory (LSTM) and Extended Long Short-Term Memory (xLSTM), with LSTM exhibiting a slight advantage in terms of Mean Squared Error (MSE). However, Spatial Long Short-Term Memory (sLSTM) and Memory-Augmented Long Short-Term Memory (mLSTM) required fewer computational resources while still achieving acceptable results. Given their lower computational cost, sLSTM and mLSTM remain viable candidates for future studies.

Figure 5 compares actual and predicted CO_2 levels, highlighting that all models remain within range, with only minor deviations at peak values.

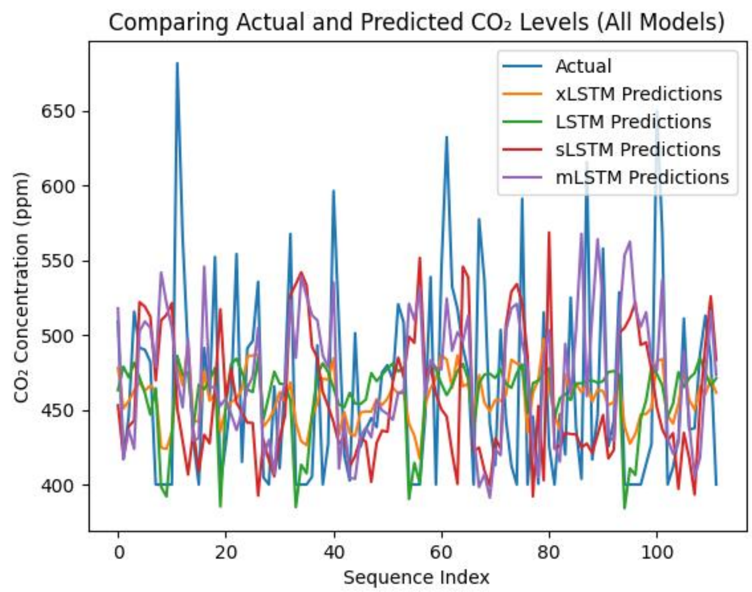

Fig. 5. Comparation of actual and predicted CO_2 levels of all models

With these different models, the results were analyzed and compared using various metrics to establish a methodology for CO_2 forecasting in indoor air quality assessments. The next section presents this comparison, highlighting the strengths and limitations of each approach.

4 Results

To ensure a comprehensive comparison, absolute errors, squared errors, logarithmic errors, percentage-based errors, and systematic bias were analyzed, providing a detailed assessment of each model's performance.

Table 2 compares the evaluation metrics for each model, highlighting differences in predictive accuracy and efficiency.

Table 2. Comparison of Evaluation Metrics for CO_2 Forecasting Models

Evaluation Metric	xLSTM	LSTM	sLSTM	mLSTM
Mean Square Error (MSE)	**3,187.804199**	3,373.60498	6,886.900879	5,951.381348
Root Mean Square Error (RMSE)	**56.460644**	58.082741	82.98735	77.145195
Mean Absolute Error (MAE)	**44.212482**	45.374516	66.055473	59.943691

Mean Absolute Logarithmic Error (MALE)	**0.113544**	0.116875	0.172207	0.160419
Mean Absolute Percentage Error (MAPE)	**9.165497**	9.413964	13.998951	13.045147
Symmetric Mean Absolute Percentage Error (SMAPE)	**9.312235**	9.512847	13.945693	12.583202
Median Absolute Error (MedAE)	**34.864059**	39.649918	54.433151	45.941559
Heber Loss	**43.712483**	44.874616	65.55547	59.447288
Theil Inequality Coefficient	**0.643643**	0.662135	0.946044	0.879444
Logarithmic Mean Squared Error (LMSE)	**0.012892**	0.01366	0.029655	0.025734
Root Logarithmic Mean Squared Error (RLMSE)	**0.113544**	0.116875	0.172207	0.160419

The evaluation of LSTM, sLSTM, mLSTM, and xLSTM models highlights key differences in their predictive accuracy and computational efficiency, with xLSTM and LSTM emerging as the most effective models.

For Mean Squared Error and Root Mean Squared Error (Willmott & Matsuura, 2005), xLSTM showed slightly lower errors than LSTM, indicating better overall precision in predicting CO_2 levels. Both significantly outperformed sLSTM and mLSTM, which exhibited considerably higher errors, making them less reliable for accurate forecasting.

Mean Absolute Error (Qi et al., 2020) results reinforced this trend, with xLSTM outperforming LSTM, indicating smaller overall deviations. Similarly, Mean Absolute Logarithmic Error demonstrated that xLSTM provided slightly better adaptability to variations in CO_2 concentration, maintaining greater accuracy across different levels.

For percentage-based errors, Mean Absolute Percentage Error (Tofallis, 2015) and Symmetric Mean Absolute Percentage Error (Kreinovich, Nguyen & Ouncharoen, 2014) indicated that xLSTM achieved lower relative errors compared to LSTM, making it more reliable in proportion-based assessments. In contrast, sLSTM and mLSTM had larger percentage-based errors, indicating reduced accuracy in capturing relative variations.

Median Absolute Error (Akinshin, 2022) further confirmed that xLSTM exhibited lower overall deviations compared to LSTM, showing a more stable error distribution.

The Huber Loss (Gokcesu & Gokcesu, 2021) results also indicated that xLSTM was slightly better at balancing squared and absolute errors, allowing for better adaptation to fluctuations in CO_2 concentration.

The Theil Inequality Coefficient (Bliemel, 1973) highlighted that xLSTM outperformed LSTM by minimizing systematic biases more effectively. sLSTM and mLSTM exhibited higher values, indicating greater prediction imbalances and less stable forecasting performance.

Finally, Logarithmic Mean Squared Error (Park, 2022) and Root Logarithmic Mean Squared Error confirmed that xLSTM had the best adaptability when predicting relative changes in CO_2 levels. LSTM followed closely, while sLSTM and mLSTM showed higher errors, indicating a reduced ability to maintain accuracy across different magnitudes.

Beyond the quantitative metrics, the observed performance differences suggest that xLSTM is capable of capturing more complex temporal features and nonlinear relationships than traditional LSTM architectures and baseline regression-based approaches. This enhanced representational capacity allows xLSTM to better model subtle variations and dependencies in indoor CO_2 dynamics, which is reflected in its consistently lower error values across all evaluated metrics.

From a learning behavior perspective, the experimental results indicate that xLSTM exhibits a more stable predictive behavior, as evidenced by smoother forecast trajectories with reduced noise compared to LSTM, sLSTM, and mLSTM. This reduced variability in the predictions suggests improved convergence stability and robustness to short-term fluctuations, particularly toward the end of the forecasting horizon, where xLSTM consistently achieves better performance than its counterparts.

Although the experiments were conducted under a single indoor environment and a stable operational phase, the consistent superiority of xLSTM across diverse error metrics provides empirical evidence of its potential generalization capability. These results can be interpreted as indicative of xLSTM's expected behavior when applied to longer temporal datasets, multiple indoor spaces, and scenarios involving more volatile CO_2 dynamics, which will be systematically explored in future work.

Overall, xLSTM consistently achieved better results than LSTM, making it the most accurate model for CO_2 forecasting. However, LSTM remained competitive, while sLSTM and mLSTM, despite their higher errors, remain viable alternatives for scenarios prioritizing computational efficiency.

5 Conclusions and Directions for Further Research

This study demonstrates that Extended Long Short-Term Memory (xLSTM) is a highly effective alternative for CO_2 forecasting in indoor environments, surpassing other LSTM-based models in predictive accuracy while maintaining computational efficiency. The results confirm that xLSTM delivers reliable and precise estimations, making it a valuable tool for monitoring air quality dynamics and supporting data-driven decision-making. Although the experimental evaluation was conducted using a limited continuous temporal subset of the original dataset and under a controlled operational phase, the consistent performance observed across the evaluated window demonstrates the model's ability to learn meaningful temporal patterns in indoor CO_2 dynamics. This controlled temporal scope provides a valid baseline for assessing model behavior while highlighting the need for future sensitivity analyses involving longer sequences and multiple non-contiguous temporal segments.

The high accuracy achieved by xLSTM is particularly relevant for policymakers and building managers who depend on precise air quality data to implement effective regulations and improvements. By leveraging advanced machine learning techniques, this study contributes to the development of healthier indoor environments, which is especially critical in urban areas where air quality management poses significant challenges. The obtained results establish an empirical reference point that supports subsequent evaluations across extended temporal horizons, diverse indoor settings, and varying operational conditions, including seasonal changes and more volatile CO_2 dynamics. The obtained metrics validate the potential of machine learning to enhance public health initiatives and support the creation of safer living and working spaces. Future work will include the integration of additional indoor air quality measurements to support more comprehensive evaluations of model robustness and generalization.

Building on these findings, future work will focus on developing a real-time system to assess respiratory disease transmission risks in indoor environments in Mexico. This system will integrate Big Data, Internet of Things (IoT), and advanced data-processing technologies using a Kappa architecture to combine real-time data streaming with batch processing for continuous air quality monitoring and analysis. It will incorporate multiple forecasting algorithms, with xLSTM as a key component for CO_2 prediction, ensuring adaptability to different indoor conditions and optimizing air quality management strategies. Future research will explicitly evaluate model sensitivity by incorporating longer temporal sequences, multiple indoor environments, and non-stationary operational phases, thereby strengthening robustness and generalization assessments. Additionally, future work will investigate combining Transformers with xLSTM models to further enhance predictive capabilities for CO_2 levels.

Further research will explore advanced analytics through deep neural networks, including image-based air quality assessment and multimodal data fusion, expanding the scope of predictive modeling. Through these efforts, this study aims to enhance indoor air quality monitoring systems and contribute to the development of safer and healthier environments across Mexico.

Data and Code Availability

The dataset and code of this study are available upon request.

Conflict of Interest

The authors declare that they have no conflicts of interest.

Ethical Considerations

This study was conducted with the utmost respect for ethical standards. Measurements were taken in a room under normal conditions during everyday activities, ensuring that subjects were not exposed to any risks. No experiments were conducted on animals or humans. Furthermore, the research activities did not produce any environmental pollution or harm. All procedures were designed to minimize any potential negative impacts and to prioritize the well-being and safety of all involved. In future work, measurements in environments with potential health risk will be performed through automated Internet of Things (IoT) technologies, using monitoring systems that operate without direct human interaction

References

Akinshin, A. (2022). *Finite-sample bias-correction factors for the median absolute deviation based on the Harrell-Davis quantile estimator and its trimmed modification*. https://doi.org/10.48550/arXiv.2207.12005

Beck, M., Pöppel, K., Spanring, M., Auer, A., Prudnikova, O., Kopp, M., Klambauer, G., Brandstetter, J., & Sepp Hochreiter (2024). *XLSTM: Extended long short-term memory*. https://doi.org/10.48550/arXiv.2405.04517

Bliemel, F. (1973). Theil's forecast accuracy coefficient: A clarification. *Journal of Marketing Research, 10*(4), 444–446. https://doi.org/10.1177/002224377301000413

Dai, Z., Yuan, Y., Zhu, X., & Zhao, L. (2024). A method for predicting indoor CO_2 concentration in university classrooms: An RF-TPE-LSTM approach. *Applied Sciences, 14*(14), 6188. https://doi.org/10.3390/app14146188

Databot. (2023). *Databot™: Real data, real science, real fun*. https://databot.us.com

Domínguez Portillo, J., Rodríguez Peralta, L. M., Sampaio, P. N. M., & Nunes, É. de O. (2023). Determining environmental indicators related to the propagation of contagious diseases and health issues: A systematic literature review. In *2023 18th Iberian Conference on Information Systems and Technologies (CISTI)* (pp. 1–5). https://doi.org/10.23919/CISTI58278.2023.10212010

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics, 29*(5), 1189–1232. https://doi.org/10.1214/aos/1013203451

Gabriel, M., & Auer, T. (2023). LSTM deep learning models for virtual sensing of indoor air pollutants: A feasible alternative to physical sensors. *Buildings, 13*(7), 1684. https://doi.org/10.3390/buildings13071684

Gabriel, M. F., Marques, G., Filipe, D., Felgueiras, F., Cardoso, J. P., Azeredo, J., Kazdaridis, G., Symeonidis, P., Keranidis, S., Conradie, P., Azevedo, I., & Anagnostopoulos, F. (2024). Implementation of an IoT architecture for promoting healthy air quality in 84 homes of families with children. *Building and Environment, 266*, 112040. https://doi.org/10.1016/j.buildenv.2024.112040

Gokcesu, K., & Gokcesu, H. (2021). *Generalized Huber loss for robust learning and its efficient minimization for robust statistics*. https://doi.org/10.48550/arXiv.2108.12627

Sepp Hochreiter, S., & Jürgen Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Inkbird. (2023). *Wi-Fi 8-in-1 air quality monitor IAQM-128W*. https://inkbird.com/products/wi-fi-air-monitor-iaqm-128w

Kreinovich, V., Nguyen, H. T., & Ouncharoen, R. (2014). *How to estimate forecasting quality: A system-motivated derivation of SMAPE*. University of Texas at El Paso. https://scholarworks.utep.edu/cs_techrep/865

Linares Alzamora, R. G., Maia Sampaio, P. N., Rodríguez Peralta, L. M., Posada Barrera, A. I., & de Oliveira Nunes, É. (2025). Sick building syndrome and indoor air quality: Leveraging Kolmogorov-Arnold networks for predictive pollutant control. In *Lecture Notes in Networks and Systems* (pp. 412–421). Springer. https://doi.org/10.1007/978-3-031-93109-3_37

Liu, Z., Wang, Y., Vaidya, S., Ruehle, F., Halverson, J., Soljačić, M., Hou, T. Y., & Max Tegmark (2024). *KAN: Kolmogorov-Arnold networks*. https://doi.org/10.48550/arXiv.2404.19756

Mohammadshirazi, A., Nadafian, A., Monsefi, A. K., Rafiei, M. H., & Ramnath, R. (2023). *Novel physics-based machine-learning models for indoor air quality approximations*. https://doi.org/10.48550/arXiv.2308.01438

Nunes, É. de O., Posada Barrera, A. I., Rodríguez Peralta, L. M., Maia Sampaio, P. N., & Cuesta Astudillo, F. L. (2025). Fuzzy risk assessment system for indoor air quality and respiratory disease prevention. *IAES International Journal of Artificial Intelligence, 14*(5), 3693–3701. https://doi.org/10.11591/ijai.v14.i5.pp3693-3701

Park, Y. (2022). *Concise logarithmic loss function for robust training of anomaly detection model*. https://doi.org/10.48550/arXiv.2201.05748

Park, J., Seo, Y., & Cho, J. (2023). Unsupervised outlier detection for time-series data of indoor air quality using LSTM autoencoder with ensemble method. *Journal of Big Data, 10*(1). https://doi.org/10.1186/s40537-023-00746-z

Posada Barrera, A. I., Rodríguez Peralta, L. M., de Oliveira Nunes, É., Maia Sampaio, P. N., & Cuesta Astudillo, F. L. (2023). Improving public health policies with indoor air quality predictive models. In *International Conference on Computational Science and Computational Intelligence (CSCI)*. IEEE. https://doi.org/10.1109/csci62032.2023.00040

Posada Barrera, A. I., Rodríguez Peralta, L. M., de Oliveira Nunes, É., Maia Sampaio, P. N., & Cuesta Astudillo, F. L. (2024a). Beyond one room: Comprehensive predictive analysis of CO_2 in indoor air quality. In *IEEE Latin American Conference on Computational Intelligence (LA-CCI)* (pp. 1–6). https://doi.org/10.1109/la-cci62337.2024.10814744

Posada Barrera, A. I., Rodríguez Peralta, L. M., de Oliveira Nunes, É., & Maia Sampaio, P. N. (2024b). Influence of indoor conditions on sick building syndrome: A data-driven investigation. In *Information Technology and Systems* (Lecture Notes in Networks and Systems, Vol. 932). Springer. https://doi.org/10.1007/978-3-031-54235-0_5

Romero-López, Y., Martínez-Cruz, A., González, R. Á., & Gálvez, A. M. S. (2025). Implementation of an IoT system with a security scheme to predict indoor CO_2 levels and mitigate COVID-19 using time series algorithms. *Integration, 103*, 102427. https://doi.org/10.1016/j.vlsi.2025.102427

Sotelo Gómez, F., Maia Sampaio, P. N., Rodríguez Peralta, L. M., Cuesta Astudillo, F. L., & Nunes, É. de O. (2025). Leveraging quantum machine learning for accurate indoor air quality forecasting and risk mitigation. In *Lecture Notes in Networks and Systems* (pp. 276–286). Springer. https://doi.org/10.1007/978-3-031-93106-2_24

Qi, J., Du, J., Siniscalchi, S. M., Ma, X., & Lee, C.-H. (2020). *On mean absolute error for deep neural network-based regression*. https://doi.org/10.48550/arXiv.2008.07281

Tofallis, C. (2015). A better measure of relative prediction accuracy for model selection and model estimation. *Journal of the Operational Research Society, 66*(8), 1352–1362. https://doi.org/10.1057/jors.2014.103

Wei, Y., Jang-Jaccard, J., Xu, W., Sabrina, F., Camtepe, S., & Boulic, M. (2023). LSTM-autoencoder-based anomaly detection for indoor air quality time-series data. *IEEE Sensors Journal, 23*(4), 3787–3800. https://doi.org/10.1109/jsen.2022.3230361

Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error over the root mean square error in assessing model performance. *Climate Research, 30*, 79–82.

Yang, G., Yuan, E., & Wu, W. (2022). Predicting long-term CO_2 concentration in classrooms based on BO–EMD–LSTM model. *Building and Environment, 224*, 109568. https://doi.org/10.1016/j.buildenv.2022.109568

Zhu, Y., Al-Ahmed, S. A., Shakir, M. Z., & Olszewska, J. I. (2022). LSTM-based IoT-enabled CO_2 forecasting for indoor air quality monitoring. *Electronics, 12*(1), 107. https://doi.org/10.3390/electronics12010107